

S/N 09/645,706



PATENT

AF/1625  
#  
JFW

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant:	Keith V. Wood et al.	Examiner:	Rebecca E. Prouty
Serial No.:	09/645,706	Group Art Unit:	1652
Filed:	August 24, 2000	Docket No.:	341.005US1
Title:	SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION		

**APPEAL BRIEF UNDER 37 C.F.R. § 41.37**

Mail Stop Appeal Brief-Patents  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is presented in support of the Notice of Appeal to the Board of Patent Appeals and Interferences, filed on March 13, 2007, from the Final Rejection of claims 1, 3-6, 9, 11-12, 15, 18, 20-21, 24-39, 41-45, 47, 60, 67, 69-71, 74, 76-78, 80-88, and 90-96 of the above-identified application, as set forth in the Final Office Action mailed on September 13, 2006.

The Commissioner of Patents and Trademarks is hereby authorized to charge Deposit Account No. 19-0743 in the amount of \$500.00 which represents the requisite fee set forth in 37 C.F.R. § 41.20(b)(2). Appellant respectfully requests consideration and reversal of the Examiner's rejection of the pending claims.

08/28/2007 SSITHIB1 00000052 190743 09645706  
02 FC:1402 500.00 DA

## TABLE OF CONTENTS

	Page
<b><u>1. REAL PARTY IN INTEREST</u></b> .....	2
<b><u>2. RELATED APPEALS AND INTERFERENCES</u></b> .....	3
<b><u>3. STATUS OF THE CLAIMS</u></b> .....	4
<b><u>4. STATUS OF AMENDMENTS</u></b> .....	5
<b><u>5. SUMMARY OF CLAIMED SUBJECT MATTER</u></b> .....	6
<b><u>6. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL</u></b> .....	16
<b><u>7. ARGUMENT</u></b> .....	17
<b><u>8. SUMMARY</u></b> .....	49
<b><u>CLAIMS APPENDIX</u></b> .....	50
<b><u>EVIDENCE APPENDIX</u></b> .....	62
<b><u>RELATED PROCEEDINGS APPENDIX</u></b> .....	64



## **1. REAL PARTY IN INTEREST**

The real party in interest of the above-captioned patent application is PROMEGA CORPORATION.

## **2. RELATED APPEALS AND INTERFERENCES**

There is an appeal in commonly assigned, copending application Serial No. 10/314,827 which may have a bearing on the Board's decision in the present appeal.

### **3. STATUS OF THE CLAIMS**

The present application was filed on August 24, 2000 with 66 claims.

Claims 1, 45, 47, and 63 were amended, claims 10, 13, 16-17, 22-23, 46, 48-59, and 65-66 were canceled, and claims 67-68 were added in the Amendment filed on August 11, 2003. Claims 1-2, 14-15, 47, 61-63, and 67-68 were amended, claims 7-8, 19 and 40 were canceled, and claims 69-73 were added in the Amendment filed on April 6, 2004. Claims 1 and 67 were amended and claims 74-80 were added in the Amendment filed on June 4, 2004.

Claims 1, 4-6, 9, 15, 18, 20-21, 24-37, 42-43, 45, 47, 60, 67, 69-71, 74, 76-78, and 80 were added, claims 2, 14, 61-63, 68, 72-73, 75, and 79 were canceled, claim 64 was withdrawn, and claims 81-82 were added in the Amendment filed on December 13, 2004. Claims 1, 47, 67, 74, and 78 were amended and claims 83-94 were added in the Amendment filed on September 22, 2005. Claims 1, 18, 44, 47, 67, 71, 74, 78, 81-85, 90, and 92 were amended, claim 89 was canceled, and claims 95-96 were added in the Amendment filed on June 19, 2006 and claims 18, 47, 83, 90, and 95-96 were amended in the Amendment filed on February 12, 2007.

Claims 1, 3-6, 9, 11-12, 15, 18, 20-21, 24-39, 41-45, 47, 60, 67, 69-71, 74, 76-78, 80-88, and 90-96 are pending and are the subject of this Appeal.

#### **4. STATUS OF AMENDMENTS**

The Rule 1.116 Final Amendment filed on February 12, 2007 was entered.

## **5. SUMMARY OF CLAIMED SUBJECT MATTER**

Some aspects of the present invention include but are not limited to synthetic nucleic acid molecules for reporter polypeptides that have nucleic acid sequences modified to remove transcription regulatory sequences.

### **Independent Claim 1**

Claim 1 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence (claims 1, 7, and 14, page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16, and page 35, line 25). The wild type nucleic acid sequence encodes chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase (page 32, lines 23-30 and page 36, line 29-page 37, line 8). The codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a reporter polypeptide which has at least 90% amino acid sequence identity to the reporter polypeptide encoded by the wild type nucleic acid sequence (page 8, lines 17-30). The codons in the second synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30, and page 65, line 21-page 66, line 3). The codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons (page 5, lines 23-26, page

8, lines 5-30, page 37, lines 16-29, page 51, lines 3-10 and 15-21, and page 51, line 28-page 52, line 5). The mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (page 37, lines 19-21).

### **Dependent Claim 90**

Claim 90, which depends on claim 1, is directed to a first synthetic nucleic acid molecule where the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the wild type nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software (page 38, lines 18-24). The identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database (page 37, lines 19-21, page 48, lines 18-24, page 48, line 29-page 49, line 6 and lines 15-18, and page 49, line 27-page 50, line 10). The codons are selected to reduce the number of identified sequences or sites (page 7, line 24-page 8, line 16 and page 48, line 17-page 52, line 11). The first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the wild type nucleic acid sequence (page 7, line 24-page 8, line 16 and page 48, line 17-page 52, line 11).

### **Independent Claim 18**

Claim 18 is directed to a synthetic nucleic acid molecule comprising SEQ ID NO:7 (GRver5), SEQ ID NO:8 (GRver6), SEQ ID NO:9 (GRver5.1), or SEQ ID NO:297 (GRver5.1), or a nucleic acid molecule which is capable of hybridizing thereto under high stringency conditions, or the complement of the hybridizable nucleic acid molecule which encodes a

luciferase. SEQ ID NOs:7, 8, 9 and 297 are synthetic nucleotide sequences of the invention encoding a "green" click beetle luciferase.

#### **Independent Claim 47**

Claim 47 is directed to a first polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297 (GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof, and comprises an open reading frame encoding a beetle luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence having SEQ ID NO:1 (claim 47). SEQ ID NO:1 is a nucleotide sequence encoding a "yellow-green" click beetle luciferase (LucPpLYG) having SEQ ID NO:23. SEQ ID NOs:18 and 301 are synthetic nucleotide sequences of the invention encoding a "red" click beetle luciferase. The codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the wild type luciferase nucleic acid sequence and is different than the codon composition of a second polynucleotide which encodes a polypeptide which has at least 90% amino acid sequence identity to the polypeptide encoded by the wild type nucleic acid sequence (page 8, lines 17-30). The codons in the second polynucleotide that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30, and page 65, line 21-page 66, line 3). The codons which differ in the first polynucleotide relative to the second polynucleotide are mammalian codons selected to result in the open reading frame in the first polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second polynucleotide by selecting the mammalian high usage codons (page 5, lines 23-26, page 8, line 5-30, page 37, lines 16-29, page 51, lines 3-10 and 15-21, and page 51, line 28-page 52, line 5). The

mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (page 37, lines 19-21).

### **Independent Claim 67**

Claim 67 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a luciferase which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type beetle luciferase nucleic acid sequence (claims 1 and 9). The codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a luciferase which has at least 90% amino acid sequence identity to the luciferase encoded by the wild type nucleic acid sequence (claims 1 and 9, and page 8, lines 17-30). The codons in the second synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30 and page 65, line 21-page 66, line 3). The codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected so as to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons (page 5, lines 23-26, page 8, lines 5-30, page 37, lines 16-29, page 51, lines 3-10 and 15-21 and page 51, line 28-page 52, line 5). The mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (page 37, lines 19-21).



**Dependent Claim 95**

Claim 95, which depends on claim 67, is directed to a first synthetic nucleic acid molecule where the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the wild type nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software (page 38, lines 18-24). The identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, and the codons are selected to reduce the number of identified sequences or sites (page 37, lines 19-21, page 48, lines 18-24, page 48, line 29-page 49, line 6 and lines 15-18 and page 48, line 17-page 50, line 10). The first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule, which has fewer mammalian transcription factor binding sequences than the wild type nucleic acid sequence (page 7, lines 24-page 8, line 16 and page 48, line 17-page 52, line 11).

**Independent Claim 74**

Claim 74 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a luciferase which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2, wherein the codon composition of the synthetic nucleic acid molecule is different at more than 25% of the codons from that of the parent nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a luciferase which has at least 90% amino acid sequence identity to the luciferase encoded by the parent nucleic acid sequence (claims 1, 7 and 14, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16 and page 35, line 25). SEQ ID NO:2 is a nucleotide sequence encoding a mutant yellow green click beetle luciferase (YG#81-6G01) (Example 1). The codons in the second synthetic nucleic acid molecule that are different than the codons in the parent nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the parent nucleic acid sequence (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30 and page 65, line 21-page 66, line 3). The codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons (page 5, lines 23-26, page 8, lines 5-30, page 37, lines 16-29, page 51, lines 3-10 and 15-21, and page 51 line 28-page 52, line 5). The mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (page 37, lines 19-21).

**Dependant Claim 96**

Claim 96, which depends on claim 74, is directed to a first synthetic nucleic acid molecule where the mammalian transcription factor binding sequences, intron splice sites,

poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the parent nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software. The identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, and the codons are selected to reduce the number of identified sequences or sites (page 37, lines 19-21, page 48, lines 18-24, page 48, line 29-page 49, line 6 and lines 15-18, and page 49, line 27-page 50, line 10). The first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the parent nucleic acid sequence (page 7, lines 24-page 8, line 16 and page 48, line 17-page 52, line 11).

### **Independent Claim 78**

Claim 78 is directed to a first polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9 (GRver5.1) or SEQ ID NO:297 (GRver5.1), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2 (claims 1, 7, 14, and 47, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16 and page 35, line 25, and Example 1). The codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the parent nucleic acid sequence and is different than the codon composition of a second polynucleotide which encodes a polypeptide which has at least 90% amino acid sequence identity to the luciferase encoded by the parent nucleic acid sequence (claims 1, 7, and 14, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16 and page 35, line 25). The codons in the second polynucleotide that are different than the codons in the parent nucleic acid sequence are mammalian high usage codons selected to result in the

second polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the parent nucleic acid sequence (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30 and page 65, line 21-page 66, line 3). The codons which differ in the first polynucleotide relative to the second polynucleotide are mammalian codons selected to result in the first polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences that are introduced to the second polynucleotide by selecting the mammalian high usage codons (page 5, lines 23-26, page 8, lines 5-30, page 37, lines 16-29, page 51, lines 3-10 and 15-21, and page 51, line 28-page 52, line 5). The mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (page 7, lines 24-page 8, line 16 and page 48, line 17-page 52, line 11).

#### **Independent Claim 83**

Claim 83 is directed to a first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297 (GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a beetle luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence (claim 47). The codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the wild type nucleic acid sequence (page 8, lines 17-30).

#### **Independent Claim 84**

Claim 84 is directed to a first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 (GRver5.1) or SEQ ID NO:297 (GRver5.1), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a polypeptide encoded by a parent

nucleic acid sequence having SEQ ID NO:2 (claim 47). The codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the parent nucleic acid sequence (page 8, lines 17-30).

### **Independent Claim 91**

Claim 91 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence, wherein the codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence (claims 1, 7, and 14, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16, and page 35, line 25). The codons in the first synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the first synthetic nucleic acid molecule having a reduced number of known mammalian transcription factor binding sequences (page 3, lines 23-26, page 4, lines 1-5 and 18-19, page 7, line 24-page 8, line 3, page 37, lines 17-30, and page 65, line 21-page 66, line 3).

### **Independent Claim 92**

Claim 92 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence (claims 1, 7, and 14, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 15-16, and page 35, line 25). The first synthetic nucleic acid molecule is prepared by replacing codons in the wild type nucleic acid molecule with mammalian high usage codons, yielding a second synthetic nucleic acid molecule, and replacing codons in the second synthetic nucleic acid molecule with mammalian codons selected to reduce the number of a combination of different, known mammalian transcription factor binding sites, yielding the first synthetic nucleic acid molecule (claims 1, 7, and 14, and page 3, lines 7-8 and 12-15, page 6, lines 6-7, page 7, lines 11-18, and page 35, line 25). The codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence (page 8, lines

16-30). The wild type nucleic acid sequence encodes chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase (page 32, lines 23-30 and page 36, line 29-page 37, line 8).

This summary does not provide an exhaustive or exclusive view of the present subject matter, and Appellant refers to the appended claims and its legal equivalents for a complete statement of the invention.

## **6. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

### **The 35 U.S.C. § 112, Second Paragraph, Rejections to be Reviewed**

Whether claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 47, 60, 67, 69-71, 74, 76-78, 80-82, 85-88, and 90-96 are unpatentable under 35 U.S.C. § 112, second paragraph.

Whether claim 90, which depends on claim 1, is unpatentable under 35 U.S.C. § 112, second paragraph.

### **The 35 U.S.C. § 112, First Paragraph, "Enablement" Rejection to be Reviewed**

Whether claims 1, 3-6, 9, 11-12, 15, 20-21, 24-33, 35-39, 41-45, 47, 60, 67, 69-70, 81-82, 86-88, and 90-95 lack enablement under 35 U.S.C. § 112, first paragraph.

### **The 35 U.S.C. § 103(a) Rejections to be Reviewed**

Whether claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 60, 67, 69-70, 81, 86, and 90-95 are unpatentable under 35 U.S.C. § 103(a) over Sherf et al. (U.S. Patent No. 5,670,356) in view of Zolotukhin et al. (U.S. Patent No. 5,874,304), Donnelly et al. (WO 97/47358), Pan et al. (Nucl. Acids Res., 27:1094 (1999)), Cornelissen et al. (U.S. Patent No. 5,952,547), and Hey et al. (U.S. Patent No. 6,169,232).

Whether claim 95, which depends on claim 67, is unpatentable under 35 U.S.C. § 103 (a) in view of Sherf et al., Zolotukhin et al., Donnelly et al., Pan et al., Cornelissen et al., and Hey et al.

Whether claims 18, 47, 71, 74, 76-78, 80, 82-85, 87-88, and 96 are unpatentable under 35 U.S.C. § 103(a) over Sherf et al., in view of Zolotukhin et al., Donnelly et al., Pan et al., Cornelissen et al., and Hey et al., and further in view of Wood et al. (WO 99/14336).

Whether claim 96, which depends on claim 74, is unpatentable under 35 U.S.C. § 103(a) over Sherf et al., Zolotukhin et al., Donnelly et al., Pan et al. Cornelissen et al., Hey et al., and further in view of Wood et al.

## **7. ARGUMENT**

### **I. The 35 U.S.C. § 112, Second Paragraph, Rejection**

#### *A) The Applicable Law under 35 U.S.C. § 112, Second Paragraph*

In rejecting a claim under the second paragraph of 35 U.S.C. § 112, it is incumbent on the Examiner to establish that one of ordinary skill in the pertinent art, when reading the claims in light of the supporting specification, would not have been able to ascertain with a reasonable degree of precision and particularity the particular area set out and circumscribed by the claims. Ex parte Wu, 10 U.S.P.Q.2d 2031, 2033 (B.P.A.I. 1989) (citing In re Moore, 439 F.2d 1232, 169 U.S.P.Q. 236 (C.C.P.A. 1971); In re Hammack, 427 F.2d 1378, 166 U.S.P.Q. 204 (C.C.P.A. 1970)).

The M.P.E.P. adopts this line of reasoning:

whether the claims set out and circumscribe a particular subject matter with a reasonable degree of clarity and particularity. Definiteness of claim language must be analyzed, not in a vacuum, but in light of: (A) The content of the particular application disclosure; (B) The teachings of the prior art; and (C) The claim interpretation that would be given by one possessing the ordinary level of skill in the pertinent art at the time the invention was made. In reviewing a claim for compliance with 35 U.S.C. 112, second paragraph, the examiner must consider the whole claim to determine whether the claim apprises one of ordinary skill in the art of its scope and, therefore, serves the notice function required by 35 U.S.C. 112, second paragraph, by providing clear warning to others as to what constitutes infringement of the patent.

M.P.E.P. § 2173.02 (emphasis added).

Moreover, if the language is as precise as the subject matter permits, the courts can demand no more. Shatterproof Glass Corp. v. Libbey-Owens Ford Co., 758 F.2d 613, 624, 225 U.S.P.Q. 634, 641 (Fed. Cir. 1985), cert. dismissed, 474 U.S. 976 (1985) (quoting Georgia-Pacific Corp. v. United States Plywood Corp., 258 F.2d 124, 136, 118 U.S.P.Q. 122, 132 (2d Cir.), cert. denied, 358 U.S. 884, 119 U.S.P.Q. 501 (1958)).



*B) The Examiner's Position*

The Examiner rejected claims 1, 3-6, 11-12, 15, 20-21, 24-39, 41-45, 47, 60, 67, 69-71, 74, 76-78, 80-82, 85-88, and 90-96 under 35 U.S.C. § 112, second paragraph, as indefinite for the recitation of "a reduced number of a combination of mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and/or prokaryotic 5' noncoding regulatory sequences", "wherein the mammalian transcription factor binding sequences are present in a database of transcription factor binding sequences" and "known mammalian transcription factor binding sequences." In particular, the Examiner asserts that the phrases at issue define a group of sequences related by function and the art does not define what sequences are included in the group, and so it would not be possible to quantify the number of such sequences.

*C) Appellant's Position*

It is Appellant's position that those skilled in the art, even in the absence of Appellant's specification, understand the metes and bounds of the phrases: transcription factor binding sequences (TFBS), intron splice sites, poly(A) addition sites, and prokaryotic 5' noncoding regulatory sequences, as they are conventionally used and understood by the art. See, e.g., U.S. Patent No. 5,670,356 ("transcription factor binding sites"), Donnelly et al. (WO 97/47358) ("intron splice sites"), Iannacone et al., *Plant Mol. Biol.*, 34:485 (1997) ("polyA sequences"), Pan et al., *Nucl. Acids Res.*, 27:1094 (1999) ("prokaryotic promoters," "poly(A) signals," and "exon-intron boundaries"), Faisst and Meyer, *Nucl. Acids Res.*, 20:3 (1992) (which discloses a compilation of vertebrate encoded transcription factors), Mount, *Am. J. Hum. Genet.*, 67:788 (2000) (consensus and other conserved splice sites), Jensen et al., *Appl. Environ. Microbiol.*, 64:82 (1998) (synthetic promoters with known consensus sequences, see abstract), Hsieh et al., *J. Bacteriol.*, 177:5740 (1995) ("a potential ribosome-binding site", see abstract and Figure 3; also see page 5742 and Figure 3 for conserved promoter sequence motifs "-10" and "-35"), and Andrews et al., *J. Virol.*, 67:7705 (1993) ("a canonical poly(A) consensus signal"; see abstract) (see Evidence Appendix). Moreover, the Examiner has acknowledged that those terms are conventional in the art (page 3 of the Office Action mailed September 13, 2006; see the Evidence Appendix).

Even assuming, for the sake of argument, the metes and bounds of the phrases TFBS, intron splice sites, poly(A) addition sites, and prokaryotic 5' noncoding regulatory sequences were not readily recognizable to the art worker, Appellant's specification discloses that a transcription regulatory element or a transcription regulatory sequence is a genetic element that controls some aspect of the expression of nucleic acid sequence(s), and includes a promoter, transcription factor binding sites, splicing signals, polyadenylation signals, termination signals, and enhancer elements (page 23, lines 24-30).

Promoters and enhancers are disclosed as typically including short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (page 24, lines 2-4). A "poly(A) sequence" is disclosed as a DNA sequence associated with the termination and polyadenylation of a nascent RNA transcript (page 25, lines 10-12). Splicing signals are disclosed as mediating the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (page 24, lines 1-2).

Moreover, the specification discloses that TFBS, intron splice sites, poly(A) addition sites, and prokaryotic 5' noncoding regulatory sequences can be identified using databases and software, e.g., databases and software such as TRANSFAC, TESS, EPD, NNPD REBASE, GenePro, MAR and BCM GeneFinder (page 38, lines 20-23). Particular TFBS, intron splice sites, poly(A) addition sites, and prokaryotic 5' noncoding regulatory sequences are shown at page 48, lines 18-24, page 49, lines 3-6 and 17, and page 50, lines 23-26 of the specification.

In addition, with regard to the phrases "wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences" and "known mammalian transcription factor binding sequences," one of skill in the art is aware of databases having transcription factor binding sequences, e.g., see page 38 and the Examples in the specification, and the Rule 132 Declaration filed on June 19, 2006 and executed by Monika Wood, a co-inventor of the above-referenced application (Evidence Appendix), or is aware of other sources of mammalian transcription factor binding sites (e.g., Faisst and Meyer, Nucl. Acids Res., 20:3 (1992); Evidence Appendix).

Further, each of the recited classes of sequences or sites has a definite property that is recognizable (and testable) by one of skill in the art. For instance, a sequence can be tested for whether it can terminate transcription and initiate poly(A) polyadenylation at the end of a RNA transcript (a poly(A) addition site); whether it can direct transcription of a gene (promoter); whether it can signal where a primary RNA transcript is to be spliced to form a mRNA (splice site); or whether it binds a transcription factor. There is nothing intrinsically wrong in using functional language, defining something by what it does rather than by what it is, in drafting patent claims; courts have even recognized the practical necessity for the use of functional language. In re Swinehart, 169 U.S.P.Q. 226, 228 (C.C.P.A. 1971).

In particular, claim 90 recites that the identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database.

Therefore, one of skill in the art in the absence of Appellant's specification or alternatively one of skill in the art in possession of Appellant's specification, would understand the metes and bounds of the phrase "mammalian transcription factor binding sequences," "intron splice sites," "poly(A) addition sites", "prokaryotic 5' noncoding regulatory sequences", "wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences" and "known mammalian transcription factor binding sequences" in the claims.

With regard to calculating the number of mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, since those sequences can be identified, as discussed above, the number present in a polynucleotide can likewise be calculated.

The Board is requested to consider that Example 1 in Appellant's specification discloses that synthetic click beetle luciferase sequences were prepared that had a reduced number of a

combination of mammalian transcription factor binding sequences, as well as intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences. In particular, it is disclosed that mammalian codon replacement in a parent click beetle luciferase sequence (YG#81-6G01) yielded a mammalian codon optimized click beetle luciferase sequence (GRver1). Removal of intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, e.g., promoter sequences, in the mammalian codon optimized click beetle luciferase sequence by codon replacement, resulted in a sequence, GRver2, that had about 100 mammalian transcription factor binding sequences. Replacement of codons in GRver2 to remove those mammalian transcription factor binding sequences, and intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, yielded a sequence, GRver3, that had about 50 newly introduced mammalian transcription factor binding sequences. Replacement of codons in GRver3 to remove those mammalian transcription factor binding sequences, and intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, yielded a sequence, GRver4, that had about 20 newly introduced mammalian transcription factor binding sequences. Those newly introduced mammalian transcription factor binding sequences were removed by codon replacement to yield GRver5.

Moreover, as described in the Rule 132 Declaration filed on June 19, 2006, using software and a database that are available to the public and comparable to those disclosed in the application, Ms. Wood determined the number of mammalian transcription factor binding sequences in *luc+*, a sequence described in Sherf et al. (U.S. Patent No. 5,670,356), a reference cited against the claims under 35 U.S.C. § 103(a).

Accordingly, the calculation of the number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences (prokaryotic promoter sequences hereinafter) in a sequence is possible and can be determined by one of skill in the art.

Regarding the Examiner's alleged change in scope of claims which recite mammalian transcription factor binding sequences, and optionally intron splice sites, poly(A) addition sites and promoter sequences, it is Appellant's position that intron splice sites, poly(A) addition sites, and prokaryotic promoter sequences represent relatively conserved sequences that were well known prior to Appellant's effective filing date (see, Mount, *supra*; Jensen et al., *supra*; Hsieh et

al., *supra*; and Andrews et al., *supra*). And although there may be new members added to the group "mammalian transcription factors" over time, the independent claims in the present application provide that the synthetic reporter nucleic acid molecules have a reduced number of a combination of different mammalian transcription factor binding sequences, as a result of codon replacement of at least 25% of the codons of a wild type or parent reporter nucleic acid sequence, with mammalian codons including mammalian high usage codons.

The reduced number of mammalian transcription factor binding sequences, and optionally intron splice sites, poly(A) addition sites and promoter sequences in Appellant's synthetic nucleic acid molecules is relative to a corresponding parent or wild type nucleic acid sequence. Wild type nucleic acid sequence for chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase were known to the art prior to Appellant's filing (Wood et al., *Science*, 244:700 (1989), Ye et al., *Biochem. Biophys. Acta*, 1339:39 (1997), Figure 2 in Murray et al., *J. Mol. Biol.*, 254:993 (1995), see Figure 4 in Zhang et al., *Proc. Natl. Acad. Sci. USA*, 94:4504 (1997), Lorenz et al., *Proc. Natl. Acad. Sci. USA*, 88:4438 (1991), and see references 1 and 21 in Sirot et al. (*Antimicrob. Agents Chemo.*, 41:1322 (1997)). Moreover, the parent reporter nucleic acid sequence recited in claims 1 and 67 is a wild type nucleic acid sequence encoding chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase, and the parent reporter nucleic acid sequence recited in claims 47, 74 and 78 has a specified wild type nucleic acid sequence (claim 47; SEQ ID NO:1) or specified parent nucleic acid sequence (claims 74 and 78; SEQ ID NO:2).

Thus, Appellant's synthetic nucleic acid molecules are readily recognized by one of skill in the art. That is because they are reporter encoding nucleotide sequences with at least 25% of codons replaced with mammalian codons relative to a corresponding wild type (or parent) reporter nucleic acid sequence, and with a reduction in a combination of different mammalian transcription factor binding sequences, as well as optionally intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, resulting from codon replacement. The presence of at least 25% mammalian codons including mammalian codons which are not "high usage" mammalian codons, and the reduction in a combination of different mammalian transcription factor binding sequences, and optionally intron splice sites, poly(A) addition sites

and prokaryotic 5' noncoding regulatory sequences, result in the synthetic nucleic acid molecules of the invention being significantly divergent in nucleotide sequence relative to the corresponding wild type or parent reporter nucleic acid sequence.

Therefore, the claims meet the requirements of 35 U.S.C. § 112, second paragraph.

## II. The 35 U.S.C. § 112, First Paragraph, "Enablement" Rejection

### *A) The Applicable Law under 35 U.S.C. § 112, First Paragraph (Enablement)*

The specification shall contain a written description of the invention and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains...to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention. 35 U.S.C. § 112(1).

It is well-settled that it is not necessary that a patent applicant have prepared and tested all the embodiments of his invention in order to meet the requirements of § 112. In re Angstadt, 190 U.S.P.Q. 214, 218 (C.C.P.A. 1976). Furthermore, enablement is not precluded by the necessity for some experimentation, such as routine screening. The key word is "undue" not "experimentation." In re Angstadt, 190 U.S.P.Q. 214, 219 (C.C.P.A. 1976). In fact, a considerable amount of experimentation is permissible if it is merely routine, or the specification provides a reasonable amount of guidance with respect to the direction in which the experimentation should take. Ex parte Jackson, 217 U.S.P.Q. 804, 807 (Bd. App. 1982).

### *B) The Examiner's Position*

The Examiner rejected claims 1, 3-6, 9, 11-12, 15, 20-21, 24-33, 35-39, 41-45, 57, 60, 67, 69-70, 81-82, 86-88, and 90-95 under 35 U.S.C. § 112, first paragraph. The Examiner asserts that the specification does not reasonably provide enablement for any variant DNA molecules encoding any reporter polypeptide having at least 90% identity to a wild type reporter polypeptide, having more than 25% of the codons altered, and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and 5' noncoding regulatory sequences than a mammalian codon optimized version of the parent

nucleic acid or to any nucleic acid which will hybridize to SEQ ID NO:9 under medium stringency conditions.

Specifically, the Examiner asserts that each of the groups of reporters that are beetle luciferases, beta-glucuronidases, chloramphenicol acetyltransferases and beta-lactamases, includes vast numbers of proteins which are not well characterized and often substantially different from those taught in the art. The Examiner further asserts that it is not routine in the art to screen for multiple substitutions or multiple modifications in proteins such as reporter proteins, and that the number of modifications encompasses many sequences, not all of which are active, and so it would require undue experimentation to make and test those modified sequences.

### *C) Appellant's Position*

One skilled in the art, having read Appellant's specification, would know how to make and use one or more synthetic nucleic acid molecules encoding a chloramphenicol acetyltransferase, beetle or *Renilla* luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase that may not be identical in amino acid sequence, but has at least 90% identity, to a reporter polypeptide encoded by a wild type or parent nucleic acid sequence for chloramphenicol acetyltransferase, beetle or *Renilla* luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase.

First, prior to the filing date of the present application, nucleotide sequences encoding, and amino acid sequences for, various reporter polypeptides were known (see, e.g., Bouthors et al., Protein Eng., 12:313 (1999); Lorenz et al. (Proc. Natl. Acad. Sci. USA, 88:4438 (1991); Matsumura et al. (Nat. Biotech., 17:696 (1999); Murray et al. (J. Mol. Biol., 254:993 (1995); Sirot et al. (Antimicrob. Agents Chemo., 41:1322 (1997); Voladri et al. (J. Bacteriol., 178:7248 (1996)), Zhang et al. (Proc. Natl. Acad. Sci. USA, 94:4504 (1997); and Wood (Science, 244:700 (1989); see Evidence Appendix). Applicant need not teach what is known to the art. Hybritech, Inc. v. Monoclonal Antibodies, Inc., 231 U.S.P.Q. 81, 94-95 (Fed. Cir. 1986).

Moreover, prior to the filing date of the present application, the relative frequency of codons employed in different organisms was known (see Aota et al., Nucl. Acids Res., 16:315 (1988), "Codon Usage Tabulated from GenBank Genetic Sequence Data"; Murray et al., Nucl.

Acids Res., 17:477 (1989), "Codon Usage in Plants"; Wada et al., Nucl. Acids Res., 18:2367 (1990) "Codon Usage Tabulated from GenBank Genetic Sequence Data"; Sharp et al., Nucl. Acids Res., 6:8207 (1988), "Codon Usage Patterns in *Escheria coli*, *Bacillus subtilus*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosiohpilia melanogaster* and Homo Sapiens; A Review of the Considerable Within Species Diversity"; and Sharp et al., Nucl. Acids Res., 15:1281 (1987), "The Codon Adaptation Index - A Measure of Directional Synonymous Codon Usage Bias, and its Potential Applications"; in Evidence Appendix). The specification also discloses codons used more frequently in human cells (page 4, lines 24-36) and codons used more frequently in plant cells (page 7, lines 1-10). Codon replacement in particular sequences is also described in U.S. Patent No. 5,670,356, Donnelly et al. (WO 97/47358), and Pan et al. (Nucl. Acids Res., 27:1094 (1999)) (see Evidence Appendix).

Further, prior to Appellant's filing, TFBS, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences were known. See Faisst and Meyer, *supra*, Mount, *supra*; Jensen et al., *supra*; Hsieh et al., *supra*; and Andrews et al., *supra*. Moreover, it was within the skill of the art to test whether a particular sequence binds transcription factors, is a splice donor or splice acceptor, is a poly(A) addition site, or initiates transcription in a prokaryotic system.

Appellant's specification describes altering the structure of a parent reporter nucleic acid sequence by iterative codon replacement to reduce TFBS, intron splice sites, poly(A) addition sites, and prokaryotic 5' noncoding regulatory sequences. Therefore, the specification, in view of the knowledge of the art worker at the time of Appellant's filing, enables synthetic reporter polypeptides of the invention regardless of the source of the parent reporter nucleic acid sequence, e.g., whether the parent nucleic acid sequence is a wild type or variant nucleic acid sequence.

Thus, a reporter polypeptide encoded by a synthetic nucleic acid molecule of the invention may include codons that result in amino acid substitutions (the synthetic nucleic acid sequence encodes a polypeptide with at least 90% amino acid sequence identity to a corresponding wild type reporter polypeptide). Appellant has provided evidence that it is well within the skill of the art to introduce substitutions into various reporter proteins and yield a variant protein with a detectable activity, e.g., an activity of the corresponding wild type reporter



protein (see, e.g., Stapleton et al., *Antimicrob. Agents Chemother.*, 43:1881 (1999); Bouthors et al., *Protein Eng.*, 12:313 (1999); Voladri et al., *J. Bacteriol.*, 178:7248 (1996); Murray et al., *J. Mol. Biol.*, 254:993 (1995); and Matsumura et al., *Nat. Biotechnol.*, 17:696 (1999); see Evidence Appendix).

In particular, with regard to luciferases, numerous substitutions have been identified in or introduced into beetle luciferases without affecting the reporter property of the substitution variants (see, e.g., Kajiyama et al., *Protein Engineering*, 4:691 (1991)), Wood et al., *J. Biolumin.*, 4:31 (1989), Wood et al., *J. Biolumin.*, 5:107 (1990) and Sala-Newby et al., *Biochem. J.*, 279:727 (1991)), U.S. Patent Nos. 5,670,356, and 6,602,677 (see Evidence Appendix). Note that LucPpyYG (SEQ ID NO:23), a wild type sequence, and YG#81-6G01 (SEQ ID NO:24) have over 95% amino acid sequence identity to each other and both function as reporters. Further, in U.S. Patent No. 6,602,677, five mutant luciferases are disclosed that have 12, 21, 32, 37 and 37 substitutions, respectively, relative to a parent luciferase, and function as reporters. Also, the Board is requested to note that in Example 1 of the above-referenced application, the amino acid sequence of the click beetle luciferases encoded by synthetic nucleic acid sequences of the invention is different than the amino acid sequence of the parent click beetle luciferase. GRver2-GRver5, and GRver5.1 have 1 amino acid substitution (related to a substitution associated with green light) relative to parent sequence YG#81-6G01; RDver2-RDver5 and RDver5.1 have 4 amino acid substitutions (related to substitutions associated with red light) relative to YG#81-6G01; RDver5.2 has 5 amino acid substitutions (related to substitutions associated with red light and improved spectral properties) relative to YG#81-6G01; and RD156-1H9 has 9 amino acid substitutions (related to substitutions associated with red light, improved spectral properties and improved luminescence intensity) relative to YG#81-6G01 (see Figure 3 for a comparison of the amino acid sequences encoded by the synthetic click beetle luciferase sequences). Similarly, in Example 3 of the present application, the amino acid sequence of the *Renilla* luciferase encoded by a synthetic nucleic acid sequence is different than the amino acid sequence of the parent *Renilla* luciferase sequence.

Moreover, the primary amino acid sequences of firefly luciferases and click beetle luciferases have common features, and so can be aligned (see Figure 8 in Wood et al., *J. Biolum. Chemi.*, 4:289 (1989), and Figure 3 in Wood et al., *Science*, 244:700 (1989); see Evidence

Appendix). Such an alignment, in view of positions that have been substituted in beetle luciferases, provides direction on what residues may be substituted without altering beetle luciferase reporter properties.

Thus, it is well within the skill of the art worker to predictably substitute amino acids in a reporter protein, e.g., substitute up to at least 10% of the residues in a luciferase, and yield a variant protein with detectable activity.

With regard to the Examiner's assertion that it is not routine to screen for multiple substitutions or multiple modifications, the Board is requested to consider WO 99/14336, Zhang et al., *supra*, and Arnold (*Chem. Eng. Sci.*, 51:5091 (1996) (see Evidence Appendix). WO 99/14336 discloses the use of recursive mutagenesis to prepare thermostable beetle luciferases, and the identification of clones with certain properties including luciferase activity. The identified clones had amino acid substitutions. Zhang et al. disclose the use of directed evolution coupled with selection to convert a beta-galactosidase to a beta-fucosidase. After iterative cycles of DNA shuffling and the evolved fucosidase screening for fucosidase activity (see Figure 13) gene encoded 8 substitutions. Arnold discloses the introduction of multiple modifications into a nucleic acid molecule and screening for particular phenotype(s) of the encoded gene product.

In response to the undue experimentation alleged to be necessary to prepare variant reporters and screen those reporters for activity, the fact that the outcome of such a screening program may be unpredictable is precisely why a screening program is carried out. It simply cannot reasonably be contended that a program to locate biomolecules with target biological or physical properties would not be carried out by the art because the results cannot be predicted in advance.

In fact, the Federal Circuit has explicitly recognized that the need, and methodologies required, to carry out extensive synthesis and screening programs to locate biomolecules with particular properties do not constitute undue experimentation. *In re Wands*, 8 U.S.P.Q.2d 1400, 1406-1407 (Fed. Cir. 1988), the Court stated:

The nature of monoclonal antibody technology is that it involves screening hybridomas to determine which ones secrete antibody with desired characteristics. Practitioners of this art are prepared to screen negative hybridomas in order to find one that makes the desired antibody.

Likewise, practitioners in the art related to the present application would be well-equipped to prepare and/or screen variant reporter constructs to identify those with reporter activity. See also, Hybritech Inc. v. Monoclonal Antibodies Inc., 231 U.S.P.Q. 81, 84 (Fed. Cir. 1986) (evidence that screening methods used to identify characteristics [of monoclonal antibodies] were available to art convincing of enablement). Thus, the fact that a given claim may encompass a variety of molecules is not dispositive of the enablement issue, particularly in an art area in which the level of skill is very high and in which screening of large numbers of compounds has been standard practice for at least ten years (Ex parte Forman, 230 U.S.P.Q.2d 456 (Bd. App. 1986)).

At page 7 in the Advisory Action date March 1, 2007, the Examiner responds that "Applicants are not claiming the screening methods they are claiming the results of the screening methods which their own argument admits are unpredictable." However, the claims at issue in In re Wands and Hybritech were not screening claims, but rather methods of using antibodies with particular properties. Further, while the outcome of a particular screening program may be unpredictable, the Federal Circuit recognized that those programs do not constitute undue experimentation. In the present application, the Board is requested to consider that screening libraries of molecules based on a parent nucleic acid molecule encoding chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase for mutated nucleic acid encoding a chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase, irrespective of whether the parent nucleic acid molecule has a wild type or variant sequence or whether the parent protein has a wild type or variant sequence, does not constitute undue experimentation, as the property to be detected ("desired characteristics") is predictable.

Claim 67 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a beetle luciferase which has at least 90% amino acid sequence identity to a beetle luciferase encoded by a wild type nucleic acid sequence. The codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a beetle luciferase which has at least 90% amino acid sequence identity to the beetle luciferase encoded by the wild type

nucleic acid sequence. The codons in the second synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence. The codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences.

Thus, even if, assuming for the sake of argument, there are vast numbers of beta-glucuronidases, chloramphenicol acetyltransferases and beta-lactamases, Appellant's specification, in view of the knowledge and skill of the art worker, enables the beetle luciferases recited in claim 67.

With respect to claim 47, which is directed to polynucleotides which hybridize to the complement of a particular synthetic nucleic acid molecule of the invention, it is Appellant's position that one of skill in the art in possession of Appellant's specification is readily able to determine whether a nucleic acid molecule hybridizes under medium stringency conditions to Appellant's synthetic polynucleotides, e.g., hybridizes to the complement of SEQ ID NO:9, and encodes a polypeptide with at least 90% amino acid sequence identity to SEQ ID NO:23. Exemplary hybridization conditions are provided at pages 20-21 in the specification.

In addition, if Appellant's specification enables a nucleic acid molecule that hybridizes under high stringency conditions to Appellant's synthetic nucleic acid molecule (note claims 71, 78-80, and 83-85 are not rejected as lacking enablement under § 112(1)), it is logical to conclude that it is within the skill of the art worker to determine whether the nucleic acid molecule also hybridizes under medium stringency conditions to Appellant's synthetic nucleic acid molecules.

Therefore, Appellant's specification fulfills the requirement of 35 U.S.C. § 112, first paragraph.

### III. The 35 U.S.C. § 103(a) Rejections

#### A) *The Applicable Law under 35 U.S.C. §103(a)*

The determination of obviousness under 35 U.S.C. § 103 is a legal conclusion based on factual evidence. *See Princeton Biochemicals, Inc. v. Beckman Coulter, Inc.*, 411 F.3d 1332, 1336-37 (Fed. Cir. 2005). The legal conclusion, that a claim is obvious within § 103(a), depends on at least four underlying factual issues set forth in *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 17, 86 S.Ct. 684, 15 L.Ed.2d 545 (1966): (1) the scope and content of the prior art; (2) differences between the prior art and the claims at issue; (3) the level of ordinary skill in the pertinent art; and (4) evaluation of any relevant secondary considerations.

The Examiner has the burden under 35 U.S.C. § 103 to establish a *prima facie* case of obviousness. *In re Fine*, 837 F.2d 1071, 5 U.S.P.Q.2d 1596 (Fed. Cir. 1988). As part of establishing a *prima facie* case of obviousness, the Examiner must show that some objective teaching in the prior art or some knowledge generally available to one of ordinary skill in the art would lead an individual to combine the relevant teaching of the references. *Id.*

The M.P.E.P. contains explicit direction that agrees with the court's holding in *In re Fine*:

In order for the Examiner to establish a *prima facie* case of obviousness, three base criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. M.P.E.P. § 2142 (citing *In re Vacek*, 947 F.2d 488, 20 U.S.P.Q.2d 1438 (Fed. Cir. 1991)).

Moreover, the Examiner must provide specific, objective evidence of record for a finding of a suggestion or motivation to combine reference teachings and must explain the reasoning by which the evidence is deemed to support such a finding. *In re Sang Su Lee*, 277 F.3d 1338, 61 U.S.P.Q.2d 1430 (Fed. Cir. 2002). Further, when making an obviousness rejection based on a

combination, there must be some motivation, suggestion or teaching of the desirability of making the specific combination made by Applicant. *Id.* Finally, the Examiner must avoid hindsight. *In re Bond*, 910 F.2d 831, 834, 15 U.S.P.Q.2d 1566, 1568 (Fed. Cir. 1990), reh'g denied, 1990 U.S. App LEXIS 1997 (Fed. Cir. 1990).

### *B) The Examiner's Position*

The Examiner rejected claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 60, 67, 69-70, 81, 86, and 89-95 under 35 U.S.C. § 103(a) as being unpatentable over Sherf et al. (U.S. Patent No. 5,670,356) in view of Zolotukhin et al. (U.S. Patent No. 5,874,304), Donnelly et al. (WO 97/47358), Pan et al. (Nucl. Acids Res., 27:1094 (1999)), Cornelissen et al. (U.S. Patent No. 5,952,547), and Hey et al. (U.S. Patent No. 6,169,232). In particular, the Examiner asserts that each of the references is drawn to methods of increasing expression in a desired host by altering the sequence of the nucleic acid but not the encoded protein in a variety of ways that will lead to increases in the protein, that the cited references show that the art was clearly aware that a combination of changes can be used to accomplish this goal, and that the art clearly teaches all of the claimed modifications and combinations of them with one or more of the others.

The Examiner also rejected claims 18, 47, 71, 74, 76-78, 80, 82-85, 87-88 and 96, under 35 U.S.C. § 103(a) as being unpatentable over Sherf et al., Donnelly et al., Zolotukhin et al., Pan et al., Cornelissen et al., Hey et al., and further in view of Wood et al. (WO 99/14336).

### *C) Appellant's Position*

#### 1. The Rejection of Claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 60, 67, 69-70, 81, 86, and 89-94

##### a. Discussion of the Cited Art

In order to minimize potential biological interferences that may complicate the interpretation of reporter data, Sherf et al. developed an optimal cytoplasmic form of luciferase (column 6, lines 7-9). Sherf et al. disclose a synthetic firefly luciferase gene (*luc*<sup>+</sup>) the sequence of which was altered primarily to remove the peroxisomal translocation sequence so as to yield a cytoplasmic form of the enzyme (abstract; column 2, lines 59-61). Other alterations were the removal of 3 internal palindromic sequences, 5 restriction endonuclease sites, 2 glycosylation

sites, and 5 to 7 transcription factor binding sites that were present in the unmodified sequence. Also codons were altered at sequences specified in Table 2 to codons preferred ("more common") in mammalian cells, relative to a wild type firefly luciferase gene (*luc*). Of the twenty 6 to 30 bp regions which were modified, 6 regions included modifications with a dual purpose, i.e., one region was modified to eliminate a glycosylation site and a transcription factor binding site that was present in the unmodified sequence, three regions were modified to eliminate a transcription factor binding site that was present in the unmodified sequence and improve codon usage, one region was modified to eliminate two transcription factor binding sites (but not improve codon usage) that were present in the unmodified sequence, and another region was modified to improve codon usage and eliminate a restriction endonuclease recognition site.

Thus, the alterations disclosed in Sherf et al. may alter post-translation steps (glycosylation), the translation product (i.e., lack of peroxisomal translocation sequence), RNA secondary structure (palindromes, and possibly RNA sequences corresponding to restriction endonuclease sites), transcription (TFBS), or sequences unrelated to transcription or translation (restriction endonuclease sites).

Sherf et al. also disclose that a vector encoding *Luc*<sup>+</sup> or *Luc* was introduced to four mammalian cell lines. NIH3T3 and HeLa cells transfected with *luc*<sup>+</sup> DNA had significantly higher levels of luciferase activity relative to NIH3T3 and HeLa cells transfected with *luc* DNA (Table 3), while CHO and CV-1 cells transfected with *luc*<sup>+</sup> or *luc* DNA had comparable luciferase activity. However, it is unclear what alterations in *luc*<sup>+</sup> DNA increased luciferase activity in mammalian cells, and why those alterations did not uniformly increase luciferase activity in all the tested mammalian cells.

In contrast, a synthetic *Renilla* luciferase gene of the present invention was expressed at significantly higher levels relative to a wild type *Renilla* luciferase gene in NIH3T3, HeLa, CHO and CV-1 cells (Table 10).

Sherf et al. do not teach or suggest that modification of a parent sequence to remove palindromic sequences, restriction endonuclease sites, glycosylation sites, and transcription factor binding sites may introduce other undesirable sequences. Nor do Sherf et al. disclose or suggest replacing at least 25% of the codons in a parent sequence with selected mammalian

codons, to reduce transcription factor binding sequences introduced to the parent sequence by high usage mammalian codons.

A humanized version of a green fluorescent protein (GFP) gene is disclosed in Zolotukhin et al. in which 88/238 of the codons in the gene were altered (column 13, lines 1-4). It is disclosed that codons were altered in order to address the poor translation efficiency of the *gfp* nucleic acid sequence in human cells (column 12, lines 49-51), as an alternative method to increase expression by, for instance, insertion of an intron (column 43, lines 26-32) or the introduction of a Kozak sequence, although insertion of a Kozak sequence did not significantly change expression (column 43, lines 49-60). Zolotukhin et al. do not disclose or suggest that codon optimization of a parent sequence may introduce undesirable sequences.

Donnelly et al. disclose the preparation of synthetic hepatitis C virus (HCV) genes for DNA vaccines. In particular, it is disclosed that codons in the corresponding wild type gene that are not the most commonly employed in humans, are replaced with an optimal codon. Of note, HCV is a pathogen of humans (page 2 of Donnelly et al.) and so due to evolutionary selection, HCV sequences are likely at least partially human codon "optimized."

Donnelly et al. also disclose that if a CG is created by that codon replacement, i.e., the third nucleotide in the replaced codon is C and the first nucleotide in the adjacent codon is G, Donnelly et al. disclose that a different codon is selected based on Table 5 in Lathe et al. (J. Mol. Biol., 183:1 (1985)) (page 17). Once all codon replacements are made, it is disclosed that the codon optimized gene is inspected for undesired sequences such as ATTTA sequences, inadvertent creation of intron splice sites, and unwanted restriction enzyme sites, which are then eliminated by substituting codons (pages 17-18).

The bias away from CG residues during codon optimization in Donnelly et al. would reduce overall CG content in the final synthetic sequence unless codon substitution to remove undesired sequences resulted in an increase in CG dinucleotides in adjacent codons (thus defeating the reasoning behind avoiding CGs in adjacent codons). In that regard, note that the synthetic click beetle and *Renilla* luciferase genes described in the Examples had increased CG content relative to the respective parent sequence.

Donnelly et al. provide no details of the sequence of any undesirable sites including intron splice sites which are to be eliminated or how to substitute codons to remove ATTTA



sequences, splicing sites and restriction enzyme sites. Further, there is no recognition in Donnelly et al. that codon optimization may introduce transcription factor binding sequences or that transcription factor binding sequences may be removed from sequences.

To address the poor expression of the merozoite surface protein-1 (*m*sp-1) gene of *Plasmodium falciparum* in heterologous systems (the wild type sequence has a high A/T content that prevented stable cloning in *E. coli* and expression in heterologous systems; abstract), Pan et al. describe a synthetic *m*sp-1 gene. Note that the life cycle of *P. falciparum* includes humans and mosquitoes (page 1094). Thus, due to evolutionary selection, *P. falciparum* sequences are likely at least partially human codon "optimized."

The synthetic gene in Pan et al. was prepared by first back translating the corresponding wild type gene using random (not preferred) human codon replacement. One master sequence was chosen with an average codon composition found in human coding sequences (page 1095), and then the master synthetic sequence was modified via alternate codon replacement to eliminate sequences that might be detrimental to efficient transcription and translation, i.e., introduce endonuclease cleavage sites to position sites at or near major processing sites for *m*sp-1, or remove prokaryotic promoters, poly(A) signals, exon-intron boundaries, prokaryotic factor-independent RNA polymerase terminators, inverted repeats (undesirable secondary structures), and long runs of purines (transcription termination) (page 1095).

Notably, Pan et al. did not seek to eliminate transcription factor binding sequences in *m*sp-1 and did not recognize that codon optimization may introduce transcription factor binding sequences. Nor does Pan et al. disclose the sequences for prokaryotic promoters, poly(A) signals, or exon-intron boundaries that are required to be identified for removal.

In the Background section of Cornelissen et al., it is disclosed that wholesale (nonselective) changes in codon usage can introduce cryptic regulatory signals in a gene, thereby causing problems in one or more of transcriptional control, RNA processing control, RNA transport control, mRNA degradation control, translational control and protein activity control, which in turn inhibits or interferes with transcription and/or translation (column 3, lines 8-25).

Nevertheless, Cornelissen et al. disclose altering DNA encoding a Bt crystal protein (insect protein) for improved expression in plants by introducing translationally neutral modification(s) in cryptic promoter(s), which can direct site-specific transcription initiation in

plant cells, and/or abortive intron(s), that inhibit or prevent transcription, nuclear accumulation of RNA or nuclear export of RNA (abstract; column 3, line 56-column 4, line 9; column 5, lines 38-45). Those modifications are the introduction of introns or replacement of codons with others encoding the same amino acid (column 3, line 56-column 4, line 9). In particular, it is disclosed that a Bt gene is modified by changing sequences with A and T to sequences with G and C that encode the same amino acids (column 8, lines 46-57; column 10, lines 14-40; column 17, lines 27-30), preferably by changing only a few nucleotides and without introducing plant preferred codons (column 10, lines 21-29). That is, to accomplish providing a translationally neutral gene, less than 10% of nucleotides in a coding region are changed from A/T to G/C (column 10, lines 16-18) and “[i]nstead of modifying the codon usage of one or more inhibitory zones” in a Bt gene, sequence elements are inactivated, preferably by introducing an intron (column 11, lines 31-35).

The Examples in Cornelissen et al. disclose methods to identify cryptic promoters and introns, and the modification of the bt884 and cryIAb22 sequences (Examples 3-4). bt884 has a 3' deletion relative to bt2 (bt2 is a wild-type gene, see Hofte et al., Eur. J. Biochem., 161:273 (1986), cited at column 17, line 29), and cryIAb22 has a 5' deletion relative to bt884 (see SEQ ID Nos. 22-23).

Cornelissen et al. do not disclose or suggest modifying reporter sequences, e.g., luciferase sequences, or recognize that codon replacement may introduce undesirable sites. Nor do Cornelissen et al. disclose or suggest replacing codons in a gene with mammalian codons.

Hey et al. disclose altering codons in storage proteins to yield sink protein nucleic acid sequences that have Trp codons for Phe codons (to increase the nutritional value of seed) and also have a reduction in splice sites, polyA sequences, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues (column 2, lines 53-67 and column 6, lines 22-25). It is disclosed that a sink protein sequence was back translated, codons preferred in maize introduced (column 10, lines 55-66), and restriction enzyme sites, splice sites, polyA sequences, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues replaced with second or third choice codons (column 11, lines 2-15).

Hey et al. do not disclose or suggest modifying reporter sequences, or recognize that mammalian codon replacement may introduce mammalian transcription factor binding sites that, in turn, may be removed.

b. The Examiner has failed to Make Out a *Prima facie* Case of Obviousness

The combination of references does not disclose or suggest Appellant's invention for the following reasons.

i. Cornelissen et al. Teach Away from the Claimed Invention and Therefore Cannot be Combined with the Other References' Teachings

Cornelissen et al. teach away from the claimed invention because Cornelissen et al. teach that very few modifications in the coding region are required, and that codon replacement with codons used more frequently in a particular host cell is not needed, to substantially alleviate expression problems. Cornelissen et al. went so far as to suggest that because only a relatively small number of modifications result in a substantial increase of foreign gene expression in plants, the modified genes produced in accordance with their invention are unlikely to contain newly introduced sequences that interfere themselves with expression of the gene in a plant cell environment (column 13, lines 37-42).

As such, it would be improper to combine Cornelissen et al. with the remaining references because Cornelissen et al. teach away from the combination. M.P.E.P. 2145 ("It is improper to combine references where the references teach away from their combination."); In re Grasselli, 713 F.2d 731, 743, 218 U.S.P.Q. 769, 779 (Fed. Cir. 1983).

ii. Impermissible Hindsight Cannot be Used to Combine the Teachings of Sherf et al. with the Other References

The combination of references does not disclose or suggest Appellant's invention as each reference discloses a different way to modify the coding sequence of a different gene to increase expression, i.e., viral genes, a gene from a parasite associated with malaria, an insect toxin gene, a storage protein gene, or a reporter gene. That is, Zolotukhin et al. disclose codon modification generally throughout a green fluorescent protein gene to codons employed more frequently in one organism, and Sherf et al. disclose limited and targeted modification (modifications in 20 regions of 6 to 30 bp) of a wild type firefly luciferase sequence to introduce or remove cloning sites, alter insect codons to mammalian codons, and to remove post-translation modification

sites, secondary structure, and transcription factor binding sites. Cornelissen et al. disclose targeted modification of a toxin gene (modify less than 10% of the nucleotides) to alter *Bacillus* codons to remove sequences that may alter elongation efficiency, e.g., by replacing A and T sequences with G and C sequences, and Donnelly et al. describe codon replacement to more commonly employed codons combined with further codon substitution to remove CG residues in adjacent codons, and subsequent inspection for ATTTA sequences, intron splice sites, and unwanted restriction enzyme sites. Hey et al. describe codon replacement in storage protein genes to maize codons and reducing restriction enzyme sites, splice sites, polyA sequences, RNA polymerase termination signals, TA and CG doublets, and blocks of G or C residues. Pan et al. disclose random human codon replacement yielding a population of synthetic sequences with codon substitutions, choosing one master synthetic sequence, and then modifying the master synthetic sequence via alternate codon replacement to eliminate sequences that might be detrimental to efficient transcription and translation, i.e., endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, exon-intron boundaries, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, and long runs of purines.

Thus, while there is a general teaching in the combination of cited documents to alter codons and/or remove certain undesired sequences in a selected sequence, none of the cited documents teaches or suggests that codon alterations, optionally in conjunction with removal of other disclosed sequences, i.e., ATTTA sequences, splice sites, endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, long runs of purines, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues, may create TFBS. The Examiner has acknowledged that none of the cited documents explicitly teaches that codon replacements may create unwanted TFBS (page 13 of the Office Action dated September 13, 2006). Moreover, none of the cited documents discloses or suggests iterative removal of TFBS from a codon altered gene of any type.

And although one of skill in the art in possession of the cited documents may be motivated to alter the codons of a particular sequence, there is no direction in the combination of cited documents which guides one of skill in the art to Appellant's invention. It is only with hindsight, i.e., given Appellant's disclosure as a "road map", that one of skill in the art, picking

and choosing from the cited documents, may be directed to Appellant's invention. That is, with regard to synthetic reporter encoding polynucleotides (claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 60, 67, 69-70, 81, 86, and 90-94), one of skill in the art in possession of the cited art, would be required to modify a reporter gene (Sherf et al. and Zolotukhin et al.), rather than a non-reporter gene (Donnelly et al., Cornelissen et al., Hey et al., and Pan et al.), by codon replacement over at least 25% of the open reading frame (Zolotukhin et al., Donnelly et al., Hey et al. and Pan et al.) rather than by alterations in a limited portion of an open reading frame (Sherf et al. and Cornelissen et al.), with subsequent additional directed alterations to the nucleotide sequence to remove undesired sequences introduced by human to optimal human, random to maize, or random human to other human, codon replacements (Donnelly et al., Hey et al. and Pan et al., respectively) rather than a lack of substantive subsequent additional directed alterations to the nucleotide sequence to remove undesired sequences introduced by codon replacement or concurrently with other alterations (Sherf et al., Zolotukhin et al., and Cornelissen et al.). In some instances, codons are further replaced to remove ATTTA sequences, splice sites, endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, long runs of purines, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues (Donnelly et al., Pan et al. and Hey et al.)

Moreover, the problem in the art (improved expression of genes in heterologous systems) has been "solved" by each of the cited documents (in different ways) and so one of skill in the art would not look to combining the references in a particular way in the absence of Appellant's disclosure.

At best, the cited documents may suggest modifying a reporter gene over a large portion of the open reading frame with a view to generally remove undesired sequences introduced by codon replacement with preferred mammalian codons, and then with other mammalian codons. But the claims at issue are not directed to such a modified reporter gene.

Assuming, for the sake of argument, that the cited documents may provide the motivation to repeat the alterations disclosed therein in a different gene, or make additional alterations in the same gene, as there is no teaching or suggestion of Appellant's invention in the cited documents taken alone or in combination, the cited documents do not provide the motivation to arrive at

Appellant's invention. That is because none of the cited documents recognizes that replacement of nonmammalian codons in a parent nucleotide sequence with mammalian codons introduces mammalian transcription factor binding sites not found in the parent nucleotide sequence. Moreover, none of the cited documents suggests that a polynucleotide that is modified by replacement of nonmammalian codons with mammalian codons be further modified by replacement with other, lower usage mammalian codons to reduce the number of introduced mammalian transcription factor binding sites.

The Board is requested to consider that after codon optimization in conjunction with removal of non-transcription factor binding sites in click beetle and *Renilla* luciferase nucleotide sequences, Appellant identified about 100 and about 60 transcription factor binding sequences, respectively. Further codon replacement to remove those sequences yielded synthetic click beetle and *Renilla* luciferase sequences with about 50 and about 20 new transcription factor binding sites, respectively, i.e., they were introduced by codon replacement (Examples 1 and 3). The vast majority of the introduced sequences were subsequently removed.

Moreover, one of skill in the art in possession of the cited documents would be required to select and identify at least transcription factor binding sites (Sherf et al. and possibly Cornelissen et al.), promoter sequences (Pan et al.), splice sites (Donnelly et al., Hey et al., and Pan et al.), and polyA sites (Hey et al. and Pan et al.) as sequences for removal by codon replacement, although the cited art would not lead the art worker to identify this specific combination of sites for alteration. Rather, Sherf et al. teach removal of peroxisomal targeting sequences, internal palindromic sequences, restriction endonuclease sites, glycosylation sites, and transcription factor binding sites, Donnelly et al. discloses removing ATTTA sequences, inadvertent creation of intron splice sites, and unwanted restriction enzyme sites, Hey et al. disclose removing splice sites, polyA sequences, RNA polymerase termination signals, TA and CG doublets, and blocks of G or C residues, and Pan et al. disclose removing endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, exon-intron boundaries, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, and long runs of purines. Cornelissen et al. modify Bt sequences by changing A and T nucleotides to G and C nucleotides. Zolotukhin et al. do not even mention removal of a set of specific regulatory sequences in a nonnative codon modified coding region.

Further, none of the cited documents discloses or suggests the use of software to identify mammalian transcription factor binding sequences in a database of transcription factor binding sequences (claim 95).

In addition, one of ordinary skill in the art in possession of the cited art would have no reasonable expectation that any particular set of changes may improve activity or be otherwise desirable in a gene that is to be expressed in a highly evolutionarily distinct cell. For instance, an increase in codon substitutions and a decrease in RNA destabilization sequences in a synthetic gene do not necessarily improve the transcriptional characteristics of the synthetic gene relative to a reference gene, as codons dictate amino acids to insert during translation and RNA destabilization sequences destabilize transcribed RNA sequences, i.e., a post-transcription process. In addition, it is unclear what changes to HCV genes (Donnelly et al.), *msh-1* gene (Pan et al.) or *luc* (Sherf et al.) sequence result in improved activity in a heterologous host and why replacement of codons in *luc* with codons preferred in mammals and other alterations which resulted in *luc*<sup>+</sup> did not improve luciferase activity in all mammalian cells which expressed Luc<sup>+</sup>.

The Examiner asserts that while it is true that none of the cited documents explicitly teach that codon replacements may create unwanted transcription factor binding sequences not present in the wild type sequence, Hey et al., Donnelly et al., and Pan et al. all show that the art recognized that codon modifications can introduce sequences which are unwanted within the synthetic gene and that additional codon modifications can decrease the introduction of those sequences, and that Sherf et al. clearly teach that the presence of transcription factor binding sequences within a reporter gene is an unwanted feature as it may interfere with the desired genetic neutrality of the reporter gene. The Examiner also asserts that it is obvious on its face that the more changes one makes, the higher the chances that such a detrimental sequence will be introduced, and that the remaining art clearly would have motivated one of skill in the art to make more substantial changes in codon preference within the luciferase of Sherf et al.

The Examiner's assertions are contradictory. Why would one of skill in the art make more changes that increase the chances that a detrimental sequence is introduced? Moreover, the template genes in Donnelly et al. and Pan et al. were already at least partially optimized for expression in a desired host, e.g., humans, as HCV and *P. falciparum* replicate and/or express

genes in humans. Further, Hey et al. teach changing the function of the protein encoded by the maize codon modified sequences.

In addition, it is likely relatively straightforward to remove a functional ATTTA sequence, splice site, restriction enzyme site, prokaryotic promoter sequence, poly(A) signal, RNA polymerase termination signals, prokaryotic factor-independent RNA polymerase terminator sequence or inverted repeat, or remove long runs of purines, TA and CG doublets, and blocks of G or C residues of more than about 4 residues (sequences disclosed as desirable to alter in Donnelly et al., Pan et al., and Hey et al.). In particular, perhaps only a single nucleotide replacement in a codon which forms part of a ATTTA sequence, intron splice site, restriction enzyme site, prokaryotic promoter, poly(A) signal, RNA polymerase termination signals, prokaryotic factor-independent RNA polymerase terminator, inverted repeat, long run of purines, TA and CG doublet, or block of G or C residues, without reference to adjacent sequences, may accomplish the removal of those undesired sequences.

In contrast, to remove a plurality of transcription factor binding sites, optionally in conjunction with other classes of sequences, by replacing codons, those modifications are selected in context, i.e., with reference to how those modifications impact adjacent sequences.

iii. The Examiner has used the Incorrect "Obvious-to-Try" Standard

In making the obviousness rejection, the Examiner clearly relies upon the discredited "obvious-to-try" standard. In re O'Farrell, 7 U.S.P.Q.2d 1673 (Fed. Cir. 1988), outlines when an invention is obvious, and therefore unpatentable, versus when an invention is obvious-to-try, and therefore patentable. The Court noted two instances in which a claimed invention is only obvious-to-try. First, an invention is merely obvious-to-try (and therefore patentable) if it is necessary to try each of numerous possible choices until one possibly arrived at a successful result, where the prior art gave no direction as to which of many possible choices is likely to be successful. 7 U.S.P.Q.2d at 1681 (citations omitted). Second, an invention is merely obvious-to-try (and therefore patentable) where the prior art gives only general guidance as to the particular form of the claimed invention or how to achieve it. At least the first of these situations applies here.

Specifically, the secondary references fail to provide any direction as to which of many possible choices of regulatory sequences in a parent nucleic acid sequence to alter but for the



specific types of sequences disclosed in those references. That is, the references fail to explicitly teach or suggest a reduction in the number of a combination of TFBS, and optionally intron splice sequences, poly(A) addition sequences, and promoter sequences. Consequently, the teachings of the secondary references fail to provide any meaningful guidance with respect to the presently claimed invention.

iv. When Combined, the References do not Teach or Suggest all the Claim Limitations

With respect to claims 1 and 67, the Board is requested to consider that none of the cited references teaches or suggests that the codons which differ in a first synthetic nucleic acid molecule relative to a second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian TFBS, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences. Those sites in the second synthetic nucleic acid molecule are the result of replacing mammalian high usage codons for codons in a wild type nucleic acid molecule encoding a reporter polypeptide, e.g., a synthetic nucleic acid molecule encoding a luciferase.

None of the cited references teach or suggest a synthetic nucleic acid molecule encoding a reporter polypeptide replacing codons in the second synthetic nucleic acid molecule with mammalian codons selected to reduce the number of known mammalian TFBS (claims 91-92). Claim 92 is directed to a first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence, where the first synthetic nucleic acid molecule is prepared by replacing codons in the wild type nucleic acid molecule with mammalian high usage codons, yielding a second synthetic nucleic acid molecule, and replacing codons in the second synthetic nucleic acid molecule with mammalian codons selected to reduce the number of a combination of different, known mammalian transcription factor binding sites, yielding the first synthetic nucleic acid molecule.

And with regard to claim 95, none of the cited references teach or suggest identifying in a wild type or second synthetic nucleic acid sequence intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, poly(A) addition sites having AATAAA,

prokaryotic 5' noncoding regulatory sequences with TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, that are identified under parameters that allow for partial ambiguity with sequences in the database.

As such, a *prima facie* case of obviousness has not been made out for claims 1, 3-6, 9, 11-12, 15, 20-21, 24-39, 41-45, 60, 67, 69-70, 81, 86, and 90-94.

## 2. The Rejection of Claims 18, 47, 71, 74, 76-78, 80, 82-85, 87-88, and 96

### a. Discussion of the Cited Art

Sherf et al., Zolotukhin et al., Donnelly et al., Pan et al., Cornelissen et al. and Hey et al. are discussed above. Wood et al. disclose thermostable beetle luciferases and a method to prepare those luciferases. It is disclosed that the thermostable beetle luciferases have a plurality of amino acid substitutions relative to a wild type beetle luciferase, and that they may be prepared by iterative mutagenesis and selection methods.

### b. The Examiner has failed to Make Out a *Prima facie* Case of Obviousness

The combination of references does not disclose or suggest Appellant's invention for the following reasons.

#### i. Cornelissen et al. Teach Away from the Claimed Invention and Therefore Cannot be Combined with the Other References' Teachings

Cornelissen et al. teach away from the claimed invention because Cornelissen et al. teach that very few modifications in the coding region are required, and that codon replacement with codons used more frequently in a particular host cell is not needed, to substantially alleviate expression problems. Cornelissen et al. went so far as to suggest that because only a relatively small number of modifications result in a substantial increase of foreign gene expression in plants, the modified genes produced in accordance with their invention are unlikely to contain newly introduced sequences that interfere themselves with expression of the gene in a plant cell environment (column 13, lines 37-42).

As such, it would be improper to combine Cornelissen et al. with the remaining references because Cornelissen et al. teach away from the combination. M.P.E.P. 2145 ("It is

improper to combine references where the references teach away from their combination."); In re Grasselli, 713 F.2d 731, 743, 218 U.S.P.Q. 769, 779 (Fed. Cir. 1983).

ii. Impermissible Hindsight Cannot be Used to Combine the Teachings of Sherf et al. with the Other References

The combination of references does not disclose or suggest Appellant's invention as each reference discloses a different way to modify the coding sequence of a different gene. While there is a general teaching in the combination of cited documents to alter codons and/or remove certain undesired sequences in a selected sequence, none of the cited documents teaches or suggests that codon alterations, optionally in conjunction with removal of ATTTA sequences, splice sites, endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, long runs of purines, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues, may create TFBS. The Examiner has acknowledged that none of the cited documents explicitly teach that codon replacements may create unwanted TFBS (page 13 of the Office Action dated September 13, 2006). Moreover, none of the cited documents discloses or suggests reiterative removal of TFBS from a codon altered gene of any type.

And although one of skill in the art in possession of the cited documents may be motivated to alter the codons of a particular sequence, there is no direction in the combination of cited documents which directs one of skill in the art to Appellant's invention. It is only with hindsight, i.e., with knowledge of Appellant's invention, that one of skill in the art, picking and choosing from the cited documents, may be directed to Appellant's invention.

With regard to Appellant's synthetic luciferase encoding polynucleotides (claims 18, 47, 71, 74, 76-78, 80, 82-85, 87-88, and 96), one of skill in the art in possession of the cited art, would be required to modify a luciferase gene (Sherf et al. and Wood et al.), rather than a non-luciferase gene (Zolotukhin et al., Donnelly et al., Cornelissen et al., Hey et al., and Pan et al.), by codon replacement over at least 25% of an open reading frame (Zolotukhin et al., Donnelly et al., Hey et al. and Pan et al.) rather than by alterations in a portion of an open reading frame (Sherf et al. and Cornelissen et al.) or by mutagenesis and selection (Wood et al.), with subsequent additional directed alterations to the nucleotide sequence to remove undesired sequences introduced by human to optimal human, random to maize, or random human to other

human, codon replacements (Donnelly et al., Hey et al. and Pan et al.) rather than a lack of substantive subsequent additional directed alterations to the nucleotide sequence to remove undesired sequences introduced by codon replacement or concurrently with other alterations, or via selection (Sherf et al., Zolotukhin et al., Cornelissen et al., and Wood et al.). In some instances, codons are further replaced to remove ATTTA sequences, splice sites, endonuclease cleavage sites, prokaryotic promoters, poly(A) signals, prokaryotic factor-independent RNA polymerase terminators, inverted repeats, long runs of purines, RNA polymerase termination signals, TA and CG doublets and blocks of G or C residues of more than about 4 residues (Donnelly et al., Pan et al. and Hey et al.).

Assuming for the sake of argument, that the cited documents may provide the motivation to repeat the alterations disclosed therein in a different gene, or make additional alterations in the same gene, as there is no teaching or suggestion of Appellant's invention in the cited documents taken alone or in combination, the cited documents do not provide the motivation to arrive at Appellant's invention. That is because none of the cited documents recognizes that replacement of nonmammalian codons in a parent polynucleotide with mammalian codons introduces mammalian transcription factor binding sites not found in the parent sequence. Moreover, none of the cited documents suggests that a polynucleotide that is modified by replacement of nonmammalian codons with mammalian codons be further modified by replacement with other, lower usage mammalian codons to reduce the number of introduced mammalian transcription factor binding sites.

iii. The Examiner has used the Incorrect "Obvious-to-Try" Standard

In making the obviousness rejection, the Examiner clearly relies upon the discredited "obvious-to-try" standard. In re O'Farrell, 7 U.S.P.Q.2d 1673 (Fed. Cir. 1988), outlines when an invention is obvious, and therefore unpatentable, versus when an invention is obvious-to-try, and therefore patentable. The Court noted two instances in which a claimed invention is only obvious-to-try. First, an invention is merely obvious-to-try (and therefore patentable) if it is necessary to try each of numerous possible choices until one possibly arrived at a successful result, where the prior art gave no direction as to which of many possible choices is likely to be successful. 7 U.S.P.Q.2d at 1681 (citations omitted). Second, an invention is merely obvious-to-try (and therefore patentable) where the prior art gives only general guidance as to the particular

form of the claimed invention or how to achieve it. At least the first of these situations applies here.

Specifically, the secondary references fail to provide any direction as to which of many possible choices of regulatory sequences in a parent nucleic acid sequence to alter but for the specific types of sequences disclosed in those references. That is, the references fail to explicitly teach or suggest a reduction in the number of a combination of TFBS, and optionally intron splice sequences, poly(A) addition sequences, and promoter sequences. Consequently, the teachings of the secondary references fail to provide any meaningful guidance with respect to the presently claimed invention.

iv. When Combined, the References do not Teach or Suggest all the Claim Limitations

With regard to claims 18, 47 and 78, none of the cited documents discloses or suggests a synthetic nucleic acid molecule comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, or SEQ ID NO:297, or a nucleic acid molecule which is capable of hybridizing thereto under high stringency conditions, or the complement of the hybridizable nucleic acid molecule which encodes a luciferase (claim 18); a first polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9, SEQ ID NO:18, SEQ ID NO:297, SEQ ID NO:301, or the complement thereof, and comprises an open reading frame encoding a beetle luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence having SEQ ID NO:1 (claim 47); or a polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9 or SEQ ID NO:297, or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2 (claim 78). In addition, none of the cited documents discloses or suggests a synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a luciferase which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2, where the codon composition of the synthetic nucleic acid molecule is different from that of the parent nucleic acid sequence and is different than the mammalian high usage codon composition of a second synthetic nucleic acid molecule which encodes a luciferase

which has at least 90% amino acid sequence identity to the luciferase encoded by the parent nucleic acid sequence, where the mammalian high usage codons are selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the parent nucleic acid sequence, and the codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of the combination that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons, and where the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences (claim 74), e.g., a synthetic luciferase encoding nucleic acid molecule where the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the parent nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software, where the identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, where the codons are selected to reduce the number of identified sequences or sites, and where the first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the parent nucleic acid sequence (claim 96).

Finally, none of the cited documents discloses or suggests a first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9, SEQ ID NO:18, SEQ ID NO:297, SEQ ID NO:301, or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to

a beetle luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence, or a first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 or SEQ ID NO:297, or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a polypeptide encoded by a parent nucleic acid sequence having SEQ ID NO:2 (claims 83 and 84).

As such, a *prima facie* case of obviousness has not been made out for claims 18, 47, 71, 74, 76-78, 80, 82-85, 87-88, and 96.

## 8. SUMMARY

It is respectfully submitted that the specification and claims of the present application satisfy the requirements of 35 U.S.C. § 112, first and second paragraphs, and that the claims are not obvious over the cited art. Therefore, reversal of the rejections and allowance of the pending claims is respectfully requested.

Respectfully submitted,

KEITH V. WOOD ET AL.

By their Representatives,

SCHWEGMAN, LUNDBERG, WOESSNER &  
KLUTH, P.A.

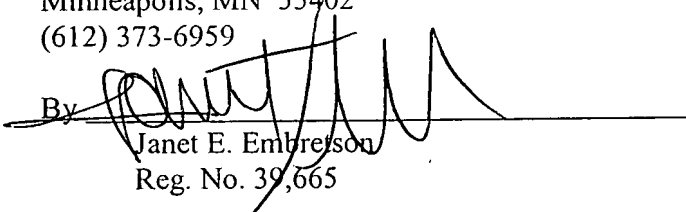
P.O. Box 2938

Minneapolis, MN 55402

(612) 373-6959

Date: August 23, 2007

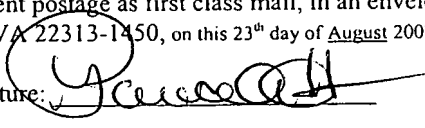
By

  
Janet E. Embreson

Reg. No. 39,665

**CERTIFICATE UNDER 37 CFR § 1.8:** The undersigned hereby certifies that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on this 23<sup>rd</sup> day of August 2007.

Name: PATRICIA A. HULTMAN

Signature: 



**CLAIMS APPENDIX**

1. A first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence, wherein the codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a reporter polypeptide which has at least 90% amino acid sequence identity to the reporter polypeptide encoded by the wild type nucleic acid sequence, wherein the codons in the second synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence, wherein the codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences, wherein the wild type nucleic acid sequence encodes chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase.

3. The first synthetic nucleic acid molecule of claim 1 wherein the codon composition of the first synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 35% of the codons.

4. The first synthetic nucleic acid molecule of claim 1 wherein the codon composition of the first synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 45% of the codons.
5. The first synthetic nucleic acid molecule of claim 1 wherein the codon composition of the first synthetic nucleic acid molecule differs from the wild type nucleic acid sequence at more than 55% of the codons.
6. The first synthetic nucleic acid molecule of claim 1 wherein the majority of codons which differ are ones that are preferred codons of a desired host cell.
9. The first synthetic nucleic acid molecule of claim 1 which encodes a luciferase.
11. The first synthetic nucleic acid molecule of claim 9 wherein the wild type nucleic acid sequence encodes a beetle luciferase.
12. The first synthetic nucleic acid molecule of claim 11 wherein the first synthetic nucleic acid molecule encodes the amino acid valine at position 224.
15. The first synthetic nucleic acid molecule of claim 1 or 9 wherein the majority of codons which differ in the second synthetic nucleic acid molecule are those which are preferred codons in humans.
18. A synthetic nucleic acid molecule comprising SEQ ID NO:7 (GRver5), SEQ ID NO:8 (GRver6), SEQ ID NO:9 (GRver5.1), or SEQ ID NO:297 (GRver5.1), or a nucleic acid molecule which is capable of hybridizing thereto under high stringency conditions, or the complement of the hybridizable nucleic acid molecule which encodes a luciferase.

20. The first synthetic nucleic acid molecule of claim 15 wherein the majority of codons which differ are the human codons CGC, CTG, TCT, AGC, ACC, CCA, CCT, GCC, GGC, GTG, ATC, ATT, AAG, AAC, CAG, CAC, GAG, GAC, TAC, TGC and TTC.
21. The first synthetic nucleic acid molecule of claim 15 wherein the majority of codons which differ are the human codons CGC, CTG, TCT, ACC, CCA, GCC, GGC, GTC, and ATC or codons CGT, TTG, AGC, ACT, CCT, GCT, GGT, GTG and ATT.
24. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule is expressed in a mammalian host cell at a level which is greater than that of the wild type nucleic acid sequence.
25. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of CTG or TTG leucine-encoding codons.
26. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of GTG or GTC valine-encoding codons.
27. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of GGC or GGT glycine-encoding codons.
28. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of ATC or ATT isoleucine-encoding codons.
29. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of CCA or CCT proline-encoding codons.
30. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of CGC or CGT arginine-encoding codons.

31. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of AGC or TCT serine-encoding codons.
32. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of ACC or ACT threonine-encoding codons.
33. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule has an increased number of GCC or GCT alanine-encoding codons.
34. The first synthetic nucleic acid molecule of claim 1 wherein the codons in the first synthetic nucleic acid molecule which differ encode the same amino acids as the corresponding codons in the wild type nucleic acid sequence.
35. A plasmid comprising the first synthetic nucleic acid molecule of claim 1.
36. An expression vector comprising the first synthetic nucleic acid molecule of claim 1 linked to a promoter functional in a cell.
37. The expression vector of claim 36 wherein the first synthetic nucleic acid molecule is operatively linked to a Kozak consensus sequence.
38. The expression vector of claim 36 wherein the promoter is functional in a mammalian cell.
39. The expression vector of claim 36 wherein the promoter is functional in a human cell.
41. The expression vector of claim 36 wherein the expression vector further comprises a multiple cloning site.

42. The expression vector of claim 41 wherein the expression vector comprises a multiple cloning site positioned between the promoter and the first synthetic nucleic acid molecule.
43. The expression vector of claim 41 wherein the expression vector comprises a multiple cloning site positioned downstream from the first synthetic nucleic acid molecule.
44. An isolated host cell comprising the expression vector of claim 36.
45. A kit comprising, in suitable container means, the expression vector of claim 36, wherein the first synthetic nucleic acid molecule encodes a reporter molecule.
47. A first polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297 (GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof, and comprises an open reading frame encoding a beetle luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence having SEQ ID NO:1, wherein the codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the wild type luciferase nucleic acid sequence and is different than the codon composition of a second polynucleotide which encodes a polypeptide which has at least 90% amino acid sequence identity to the polypeptide encoded by the wild type nucleic acid sequence, wherein the codons in the second polynucleotide that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence, wherein the codons which differ in the first polynucleotide relative to the second polynucleotide are mammalian codons selected to result in the open reading frame in the first polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding

regulatory sequences, that are introduced to the second polynucleotide by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences.

60. The first synthetic nucleic acid molecule of claim 1 wherein the first synthetic nucleic acid molecule is expressed at a level which is at least 110% of that of the wild type nucleic acid sequence in a cell or cell extract under identical conditions.

67. A first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a luciferase which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type beetle luciferase nucleic acid sequence, wherein the codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a luciferase which has at least 90% amino acid sequence identity to the luciferase encoded by the wild type nucleic acid sequence, wherein the codons in the second synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the wild type nucleic acid sequence, wherein the codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected so as to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences.

69. The first synthetic nucleic acid molecule of claim 11 or 67 which has 74% or less nucleic acid sequence identity to the wild type nucleic acid sequence.

70. The first synthetic nucleic acid molecule of claim 11 or 67 which has at least 40-fold increased expression relative to the wild type nucleic acid sequence.

71. The first polynucleotide of claim 47 which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297(GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof.

74. A first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a luciferase which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2, wherein the codon composition of the synthetic nucleic acid molecule is different at more than 25% of the codons from that of the parent nucleic acid sequence and is different than the codon composition of a second synthetic nucleic acid molecule which encodes a luciferase which has at least 90% amino acid sequence identity to the luciferase encoded by the parent nucleic acid sequence, wherein the codons in the second synthetic nucleic acid molecule that are different than the codons in the parent nucleic acid sequence are mammalian high usage codons selected to result in the second synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the parent nucleic acid sequence, wherein the codons which differ in the first synthetic nucleic acid molecule relative to the second synthetic nucleic acid molecule are mammalian codons selected to result in the first synthetic nucleic acid molecule having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences, that are introduced to the second synthetic nucleic acid molecule by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences.

76. The first synthetic nucleic acid molecule of claim 74 wherein the polypeptide encoded by the first synthetic nucleic acid molecule has at least 95% amino acid identity to the luciferase encoded by the parent nucleic acid sequence.

77. The first synthetic nucleic acid molecule of claim 74 which has 74% or less nucleic acid sequence identity to the parent nucleic acid sequence.

78. A first polynucleotide which hybridizes under medium stringency hybridization conditions to SEQ ID NO:9 (GRver5.1) or SEQ ID NO:297 (GRver5.1), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase encoded by a parent nucleic acid sequence having SEQ ID NO:2, wherein the codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the parent nucleic acid sequence and is different than the codon composition of a second polynucleotide which encodes a polypeptide which has at least 90% amino acid sequence identity to the luciferase encoded by the parent nucleic acid sequence, wherein the codons in the second polynucleotide that are different than the codons in the parent nucleic acid sequence are mammalian high usage codons selected to result in the second polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences relative to the parent nucleic acid sequence, wherein the codons which differ in the first polynucleotide relative to the second polynucleotide are mammalian codons selected to result in the first polynucleotide having a reduced number of a combination of different mammalian transcription factor binding sequences, and optionally a reduced number of intron splice sites, poly(A) addition sites or prokaryotic 5' noncoding regulatory sequences that are introduced to the second polynucleotide by selecting the mammalian high usage codons, wherein the mammalian transcription factor binding sequences are those present in a database of transcription factor binding sequences.



80. The first polynucleotide of claim 78 wherein the polypeptide encoded by the first polynucleotide has at least 95% amino acid identity to the luciferase encoded by the parent nucleic acid molecule.

81. The first synthetic nucleic acid molecule of claim 1, 67 or 74 wherein the transcription factor binding sequence is at least 5 bases in length.

82. The first polynucleotide of claim 47 or 78 wherein the transcription factor binding sequence is at least 5 bases in length.

83. A first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 (GRver5.1), SEQ ID NO:18 (RD156-1H9), SEQ ID NO:297 (GRver5.1), SEQ ID NO:301 (RD156-1H9), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a beetle luciferase having SEQ ID NO:23 encoded by a corresponding wild type nucleic acid sequence, wherein the codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the wild type nucleic acid sequence.

84. A first polynucleotide which hybridizes under high stringency hybridization conditions to SEQ ID NO:9 (GRver5.1) or SEQ ID NO:297 (GRver5.1), or the complement thereof, and comprises an open reading frame encoding a luciferase polypeptide which has at least 90% amino acid sequence identity to a polypeptide encoded by a parent nucleic acid sequence having SEQ ID NO:2, wherein the codon composition of the open reading frame of the first polynucleotide is different at more than 25% of the codons from that of the parent nucleic acid sequence.

85. The first polynucleotide of claim 78 which hybridizes under high stringency conditions.

86. The first synthetic sequence of claim 1 wherein the selection of mammalian high usage codons and mammalian codons also reduces the number of restriction endonuclease sites.

87. The first polynucleotide of claim 47 or 78 wherein the selection of mammalian high usage codons and mammalian codons also reduces the number of restriction endonuclease sites.

88. The first synthetic nucleic acid molecule of claim 67 or 74 wherein the selection of mammalian high usage codons and mammalian codons also reduces the number of restriction endonuclease sites.

90. The first synthetic nucleic acid molecule of claim 1 wherein the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the wild type nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software, wherein the identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, wherein the codons are selected to reduce the number of identified sequences or sites, and wherein the first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the wild type nucleic acid sequence.

91. A first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence, wherein the codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence, wherein the codons in the first synthetic nucleic acid molecule that are different than the codons in the wild type nucleic acid sequence

are mammalian high usage codons selected to result in the first synthetic nucleic acid molecule having a reduced number of known mammalian transcription factor binding sequences.

92. A first synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a reporter polypeptide which has at least 90% amino acid sequence identity to a reporter polypeptide encoded by a wild type nucleic acid sequence, wherein the first synthetic nucleic acid molecule is prepared by replacing codons in the wild type nucleic acid molecule with mammalian high usage codons, yielding a second synthetic nucleic acid molecule, and replacing codons in the second synthetic nucleic acid molecule with mammalian codons selected to reduce the number of a combination of different, known mammalian transcription factor binding sites, yielding the first synthetic nucleic acid molecule, wherein the codon composition of the first synthetic nucleic acid molecule is different at more than 25% of the codons from that of the wild type nucleic acid sequence, wherein the wild type nucleic acid sequence encodes chloramphenicol acetyltransferase, *Renilla* luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase.

93. The first synthetic nucleic acid molecule of claim 91 or 92 which has at least 2-fold fewer mammalian transcription factor binding sequences relative to the wild type nucleic acid sequence.

94. The first synthetic nucleic acid molecule of claim 91 or 92 wherein codon selection also reduces the number of intron splice sites, poly(A) addition sites or promoter sequences, or a combination thereof.

95. The first synthetic nucleic acid molecule of claim 67 wherein the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the wild type nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software, wherein the identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are

selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, wherein the codons are selected to reduce the number of identified sequences or sites, and wherein the first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the wild type nucleic acid sequence.

96. The first synthetic nucleic acid molecule of claim 74 wherein the mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in the parent nucleic acid sequence or the second synthetic nucleic acid sequence are identified with software, wherein the identified intron splice sites are selected from AGGTRAGT, AGGTRAG, GGTRAGT or YNCAGG, the identified poly(A) addition sites have AATAAA, the identified prokaryotic 5' noncoding regulatory sequences are selected from TATAAT, or AGGA or GGAG if a methionine codon is within 12 bases 3' of the AGGA or GGAG, and the identified mammalian transcription factor binding sequences are in a database of transcription factor binding sequences, mutant transcription factor binding sequences and consensus transcription factor binding sequences, and identified under parameters that allow for partial ambiguity with sequences in the database, wherein the codons are selected to reduce the number of identified sequences or sites, and wherein the first synthetic nucleic acid molecule has fewer mammalian transcription factor binding sequences than the second synthetic nucleic acid molecule which has fewer mammalian transcription factor binding sequences than the parent nucleic acid sequence.

## **EVIDENCE APPENDIX**

U.S. Patent No. 5,670,356; cited on Form 1449 filed on June 13, 2001

U.S. Patent No. 6,602,677; cited on Form 1449 filed on June 13, 2001

WO 97/47358; cited on Form 892 which accompanied the Office Action dated September 13, 2004

WO 99/14336; cited on Form 1449 filed on June 13, 2001

Andrews et al., J. Virol., 67:7705 (1993); cited on Form 1449 filed on July 6, 2006

Aota et al., Nucl. Acids Res., 16:315 (1988); cited on Form 1449 filed on June 13, 2001

Arnold (Chem. Eng. Sci., 51:5091 (1996); cited on Form 1449 filed on July 6, 2006

Bouthors et al., Protein Eng., 12:313 (1999); cited on Form 1449 filed on July 10, 2006

Faisst and Meyer, Nucl. Acids Res., 20:3 (1992); cited on Form 1449 filed on June 13, 2001

Hsieh et al., J. Bact., 177:5740 (1995); cited on Form 1449 filed on July 6, 2006

Jensen et al., Appl. Environ. Microbiol., 64:82 (1998); cited on Form 1449 filed on July 6, 2006

Kajiyama et al., Protein Engineering, 4:691 (1991); cited on Form 1449 filed on July 6, 2006

Lorenz et al., Proc. Natl. Acad. Sci. USA, 88:4438 (1991); cited on Form 1449 filed on July 10, 2006

Matsumura et al., Nat. Biotech., 17:696 (1999); cited on Form 1449 filed on July 10, 2006

Mount, Am. J. Hum. Genet., 67:788 (2000); cited on Form 1449 filed on July 10, 2006

Murray et al., Nucl. Acids Res., 17:477 (1989); cited on Form 1449 filed on June 13, 2001

Murray et al., J. Mol. Biol., 254:993 (1995); cited on Form 1449 filed on July 10, 2006

Pan et al., Nucl. Acids Res., 27:1094 (1999); cited on Form 1449 filed on August 11, 2003

Sala-Newby et al., Biochem. J., 279:727 (1991); cited on Form 1449 filed on July 6, 2006

Sharp et al., Nucl. Acids Res., 15:1281 (1987); cited on Form 1449 filed on July 6, 2006

Sharp et al., Nucl. Acids Res., 16:8207 (1988); cited on Form 1449 filed on July 6, 2006

Sirot et al., Antimicrob. Agents Chemo., 41:1322 (1997); cited on Form 1449 filed on July 10, 2006

Stapleton et al., Antimicrob. Agents Chemother., 43:1881 (1999); cited on Form 1449 filed on July 6, 2006

Voladri et al., J. Bacteriol., 178:7248 (1996); cited on Form 1449 filed on July 10, 2006

Wada et al., Nucl. Acids Res., 18:2367 (1990); cited on Form 1449 filed on June 13, 2006

Wood et al., J. Biolumin., 4:31 (1989); cited on Form 1449 filed on July 6, 2006

Wood et al., J. Biolum. Chemi., 4:289 (1989); cited on Form 1449 filed on July 6, 2006

Wood et al., J. Biolumin., 5:107 (1990); cited on Form 1449 filed on July 6, 2006

Wood, Science, 244:700 (1989); cited on Form 1449 filed on July 10, 2006

Ye et al., Bioch. Biophys. Acta, 1339:39 (1997); cited on Form 1449 filed on July 6, 2006

Zhang et al., Proc. Natl. Acad. Sci. USA, 94:4504 (1997); cited on Form 1449 filed July 10, 2006

Office Action dated September 13, 2006

Wood Rule 132 Declaration filed on June 19, 2006

**RELATED PROCEEDINGS APPENDIX**

None.

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant: Keith V. Wood et al.

Title: SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION

Docket No.: 341.005US1

Filed: August 24, 2000

Examiner: Rebecca E. Prouty

Serial No.: 09/645,706

Due Date: September 13, 2007

Group Art Unit: 1652



**MS Appeal Brief - Patents**

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

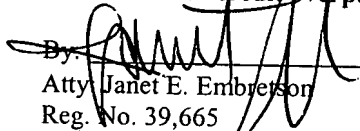
We are transmitting herewith the following attached items (as indicated with an "X"):

- X Return postcards (Postcards 1 of 3; 2 of 3 and 3 of 3)
- X Appeal Brief Under 37 CFR 41.37 (64 pgs.)
- X Petition for Extension of Time (1 pg.)
- X Authorization to charge Deposit Account No. 190743 in the amount of \$1590.00 to cover the Extension of Time Fee.
- X Authorization to charge Deposit Account No. 190743 in the amount of \$500.00 to cover the Appeal Brief Fee.
- X U.S. Patent No. 5,670,356; cited on Form 1449 filed on June 13, 2001
- X U.S. Patent No. 6,602,677; cited on Form 1449 filed on June 13, 2001
- X WO 97/47358; cited on Form 892 which accompanied the Office Action dated September 13, 2004
- X WO 99/14336; cited on Form 1449 filed on June 13, 2001
- X Andrews et al., J. Virol., 67:7705 (1993); cited on Form 1449 filed on July 6, 2006
- X Aota et al., Nucl. Acids Res., 16:315 (1988); cited on Form 1449 filed on June 13, 2001
- X Arnold (Chem. Eng. Sci., 51:5091 (1996); cited on Form 1449 filed on July 6, 2006
- X Bouthors et al., Protein Eng., 12:313 (1999); cited on Form 1449 filed on July 10, 2006
- X Faisst and Meyer, Nucl. Acids Res., 20:3 (1992); cited on Form 1449 filed on June 13, 2001
- X Hsieh et al., J. Bact., 177:5740 (1995); cited on Form 1449 filed on July 6, 2006
- X Jensen et al., Appl. Environ. Microbiol., 64:82 (1998); cited on Form 1449 filed on July 6, 2006
- X Kajiyama et al., Protein Engineering, 4:691 (1991); cited on Form 1449 filed on July 6, 2006
- X Lorenz et al., Proc. Natl. Acad. Sci. USA, 88:4438 (1991); cited on Form 1449 filed on July 10, 2006
- X Matsumura et al., Nat. Biotech., 17:696 (1999); cited on Form 1449 filed on July 10, 2006
- X Mount, Am. J. Hum. Genet., 67:788 (2000); cited on Form 1449 filed on July 10, 2006
- X Murray et al., Nucl. Acids Res., 17:477 (1989); cited on Form 1449 filed on June 13, 2001
- X Murray et al., J. Mol. Biol., 254:993 (1995); cited on Form 1449 filed on July 10, 2006
- X Pan et al., Nucl. Acids Res., 27:1094 (1999); cited on Form 1449 filed on August 11, 2003
- X Sala-Newby et al., Biochem. J., 279:727 (1991); cited on Form 1449 filed on July 6, 2006
- X Sharp et al., Nucl. Acids Res., 15:1281 (1987); cited on Form 1449 filed on July 6, 2006
- X Sharp et al., Nucl. Acids Res., 16:8207 (1988); cited on Form 1449 filed on July 6, 2006
- X Sirot et al., Antimicrob. Agents Chemo., 41:1322 (1997); cited on Form 1449 filed on July 10, 2006
- X Stapleton et al., Antimicrob. Agents Chemother., 43:1881 (1999); cited on Form 1449 filed on July 6, 2006
- X Voladri et al., J. Bacteriol., 178:7248 (1996); cited on Form 1449 filed on July 10, 2006
- X Wada et al., Nucl. Acids Res., 18:2367 (1990); cited on Form 1449 filed on June 13, 2006
- X Wood et al., J. Biolumin., 4:31 (1989); cited on Form 1449 filed on July 6, 2006
- X Wood et al., J. Biolum. Chemi., 4:289 (1989); cited on Form 1449 filed on July 6, 2006
- X Wood et al., J. Biolumin., 5:107 (1990); cited on Form 1449 filed on July 6, 2006
- X Wood, Science, 244:700 (1989); cited on Form 1449 filed on July 10, 2006
- X Ye et al., Bioch. Biophys. Acta, 1339:39 (1997); cited on Form 1449 filed on July 6, 2006
- X Zhang et al., Proc. Natl. Acad. Sci. USA, 94:4504 (1997); cited on Form 1449 filed July 10, 2006
- X Office Action dated September 13, 2006
- X Wood Rule 132 Declaration filed on June 19, 2006

**If not provided for in a separate paper filed herewith, Please consider this a PETITION FOR EXTENSION OF TIME for sufficient number of months to enter these papers and please charge any additional fees or credit overpayment to Deposit Account No. 19-0743.**

**SCHWEGMAN, LUNDBERG & WOESSNER, P.A.**

Customer Number 21186

By:   
Atty Janet E. Embretson  
Reg. No. 39,665

CERTIFICATE UNDER 37 CFR 1.8: The undersigned hereby certifies that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail, in an envelope addressed to: MS Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on this 23<sup>rd</sup> day of August, 2007.

PATRICIA A. HULTMAN

Name

SCHWEGMAN, LUNDBERG & WOESSNER, P.A.

  
Signature

(GENERAL)



**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>A61N 43/04, C12Q 1/68, C12N 15/00, C07H 21/02, A61K 39/00</b>	<b>A1</b>	<b>(11) International Publication Number: WO 97/47358</b> <b>(43) International Publication Date: 18 December 1997 (18.12.97)</b>
<b>(21) International Application Number:</b> PCT/US97/09884 <b>(22) International Filing Date:</b> 6 June 1997 (06.06.97)  <b>(30) Priority Data:</b> 60/020,494 11 June 1996 (11.06.96) US 9614731.9 12 July 1996 (12.07.96) GB 60/033,534 20 December 1996 (20.12.96) US  <b>(71) Applicant (for all designated States except US):</b> MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DONNELLY, John, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LIU, Margaret, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).  <b>(74) Common Representative:</b> MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).		<b>(81) Designated States:</b> AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> SYNTHETIC HEPATITIS C GENES  <b>(57) Abstract</b>  This invention relates to novel methods and formulations of nucleic acid pharmaceutical products, specifically formulations of nucleic acid vaccine products and nucleic acid gene therapy products.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

TITLE OF THE INVENTION  
SYNTHETIC HEPATITIS C GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

5 Not applicable.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D  
Not applicable.

10 REFERENCE TO MICROFICHE APPENDIX  
Not applicable.

FIELD OF THE INVENTION  
Not applicable.

15 BACKGROUND OF THE INVENTION

This invention relates to novel nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the  
20 production of immune responses which specifically recognize Hepatitis C virus (HCV).

Hepatitis C Virus

Non-A, Non-B hepatitis (NANBH) is a transmissible disease  
25 (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three  
30 types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH).

- 2 -

Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma. Thus, a need exists for an effective method  
5 for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in M. A. Brinton, in  
10 "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm.  
15 Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000  
20 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded,  
25 and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known  
30 viral sequences, small but significant co-linear homologs are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

- 3 -

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the in situ generation of the protein in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were  
5 generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result  
10 in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means  
15 offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

Therefore, this invention contemplates methods for introducing nucleic acids into living tissue to induce expression of proteins. The invention provides a method for introducing viral  
20 proteins into the antigen processing pathway to generate virus-specific immune responses including, but not limited to, CTLs. Thus, the need for specific therapeutic agents capable of eliciting desired prophylactic immune responses against viral pathogens is met for HCV virus by this invention. Of particular importance in this therapeutic approach is the  
25 ability to induce T-cell immune responses which can prevent infections even of virus strains which are heterologous to the strain from which the antigen gene was obtained. Therefore, this invention provides DNA constructs encoding viral proteins of the hepatitis C virus core, envelope (E1), nonstructural (NS5) genes or any other HCV genes which encode  
30 products which generate specific immune responses including but not limited to CTLs.

- 4 -

### DNA Vaccines

- Benvenisty, N., and Reshef, L. [PNAS 83, 9551-9555, (1986)] showed that  $\text{CaCl}_2$ -precipitated DNA introduced into mice intraperitoneally (i.p.), intravenously (i.v.) or intramuscularly (i.m.) could be expressed. The i.m. injection of DNA expression vectors without  $\text{CaCl}_2$  treatment in mice resulted in the uptake of DNA by the muscle cells and expression of the protein encoded by the DNA. The plasmids were maintained episomally and did not replicate. Subsequently, persistent expression has been observed after i.m. injection in skeletal muscle of rats, fish and primates, and cardiac muscle of rats. The technique of using nucleic acids as therapeutic agents was reported in WO90/11092 (4 October 1990), in which polynucleotides were used to vaccinate vertebrates.
- It is not necessary for the success of the method that immunization be intramuscular. The introduction of gold microprojectiles coated with DNA encoding bovine growth hormone (BGH) into the skin of mice resulted in production of anti-BGH antibodies in the mice. A jet injector has been used to transfect skin, muscle, fat, and mammary tissues of living animals. Various methods for introducing nucleic acids have been reviewed. Intravenous injection of a DNA:cationic liposome complex in mice was shown by Zhu et al., [Science 261:209-211 (9 July 1993)] to result in systemic expression of a cloned transgene. Ulmer et al., [Science 259:1745-1749, (1993)] reported on the heterologous protection against influenza virus infection by intramuscular injection of DNA encoding influenza virus proteins.
- The need for specific therapeutic and prophylactic agents capable of eliciting desired immune responses against pathogens and tumor antigens is met by the instant invention. Of particular importance in this therapeutic approach is the ability to induce T-cell immune responses which can prevent infections or disease caused even by virus strains which are heterologous to the strain from which the antigen gene was obtained. This is of particular concern when dealing with HIV as this virus has been recognized to mutate rapidly and many virulent isolates have been identified [see, for example, LaRosa et al.,

- 5 -

Science 249:932-935 (1990), identifying 245 separate HIV isolates]. In response to this recognized diversity, researchers have attempted to generate CTLs based on peptide immunization. Thus, Takahashi et al., [Science 255:333-336 (1992)] reported on the induction of broadly  
5 cross-reactive cytotoxic T cells recognizing an HIV envelope (gp160) determinant. However, those workers recognized the difficulty in achieving a truly cross-reactive CTL response and suggested that there is a dichotomy between the priming or restimulation of T cells, which is very stringent, and the elicitation of effector function, including  
10 cytotoxicity, from already stimulated CTLs.

Wang et al. reported on elicitation of immune responses in mice against HIV by intramuscular inoculation with a cloned, genomic (unspliced) HIV gene. However, the level of immune responses achieved in these studies was very low. In addition, the Wang et al.,  
15 DNA construct utilized an essentially genomic piece of HIV encoding contiguous Tat/REV-gp160-Tat/REV coding sequences. As is described in detail below, this is a suboptimal system for obtaining high-level expression of the gp160. It also is potentially dangerous because expression of Tat contributes to the progression of Kaposi's Sarcoma.

20 WO 93/17706 describes a method for vaccinating an animal against a virus, wherein carrier particles were coated with a gene construct and the coated particles are accelerated into cells of an animal.

The instant invention contemplates any of the known methods for introducing polynucleotides into living tissue to induce  
25 expression of proteins. However, this invention provides a novel immunogen for introducing proteins into the antigen processing pathway to efficiently generate specific CTLs and antibodies.

#### Codon Usage and Codon Context

30 The codon pairings of organisms are highly nonrandom, and differ from organism to organism. This information is used to construct and express altered or synthetic genes having desired levels of translational efficiency, to determine which regions in a genome are protein coding regions, to introduce translational pause sites into

- 6 -

heterologous genes, and to ascertain relationship or ancestral origin of nucleotide sequences

The expression of foreign heterologous genes in transformed organisms is now commonplace. A large number of mammalian genes, including, for example, murine and human genes, have been successfully inserted into single celled organisms. Standard techniques in this regard include introduction of the foreign gene to be expressed into a vector such as a plasmid or a phage and utilizing that vector to insert the gene into an organism. The native promoters for such genes are commonly replaced with strong promoters compatible with the host into which the gene is inserted. Protein sequencing machinery permits elucidation of the amino acid sequences of even minute quantities of native protein. From these amino acid sequences, DNA sequences coding for those proteins can be inferred. DNA synthesis is also a rapidly developing art, and synthetic genes corresponding to those inferred DNA sequences can be readily constructed.

Despite the burgeoning knowledge of expression systems and recombinant DNA, significant obstacles remain when one attempts to express a foreign or synthetic gene in an organism. Many native, active proteins, for example, are glycosylated in a manner different from that which occurs when they are expressed in a foreign host. For this reason, eukaryotic hosts such as yeast may be preferred to bacterial hosts for expressing many mammalian genes. The glycosylation problem is the subject of continuing research.

Another problem is more poorly understood. Often translation of a synthetic gene, even when coupled with a strong promoter, proceeds much less efficiently than would be expected. The same is frequently true of exogenous genes foreign to the expression organism. Even when the gene is transcribed in a sufficiently efficient manner that recoverable quantities of the translation product are produced, the protein is often inactive or otherwise different in properties from the native protein.



- 7 -

It is recognized that the latter problem is commonly due to differences in protein folding in various organisms. The solution to this problem has been elusive, and the mechanisms controlling protein folding are poorly understood.

5           The problems related to translational efficiency are believed to be related to codon context effects. The protein coding regions of genes in all organisms are subject to a wide variety of functional constraints, some of which depend on the requirement for encoding a properly functioning protein, as well as appropriate  
10 translational start and stop signals. However, several features of protein coding regions have been discerned which are not readily understood in terms of these constraints. Two important classes of such features are those involving codon usage and codon context.

          It is known that codon utilization is highly biased and varies  
15 considerably between different organisms. Codon usage patterns have been shown to be related to the relative abundance of tRNA isoacceptors. Genes encoding proteins of high versus low abundance show differences in their codon preferences. The possibility that biases in codon usage alter peptide elongation rates has been widely discussed.  
20 While differences in codon use are associated with differences in translation rates, direct effects of codon choice on translation have been difficult to demonstrate. Other proposed constraints on codon usage patterns include maximizing the fidelity of translation and optimizing the kinetic efficiency of protein synthesis.

25           Apart from the non-random use of codons, considerable evidence has accumulated that codon/anticodon recognition is influenced by sequences outside the codon itself, a phenomenon termed "codon context." There exists a strong influence of nearby nucleotides on the efficiency of suppression of nonsense codons as well as missense codons.  
30 Clearly, the abundance of suppressor activity in natural bacterial populations, as well as the use of "termination" codons to encode selenocysteine and phosphoserine require that termination be context-dependent. Similar context effects have been shown to influence the fidelity of translation, as well as the efficiency of translation initiation.

- 8 -

Statistical analyses of protein coding regions of E. coli have demonstrate another manifestation of "codon context." The presence of a particular codon at one position strongly influences the frequency of occurrence of certain nucleotides in neighboring codons, and these  
5 context constraints differ markedly for genes expressed at high versus low levels. Although the context effect has been recognized, the predictive value of the statistical rules relating to preferred nucleotides adjacent to codons is relatively low. This has limited the utility of such nucleotide preference data for selecting codons to effect desired levels  
10 of translational efficiency.

The advent of automated nucleotide sequencing equipment has made available large quantities of sequence data for a wide variety of organisms. Understanding those data presents substantial difficulties. For example, it is important to identify the coding regions of the  
15 genome in order to relate the genetic sequence data to protein sequences. In addition, the ancestry of the genome of certain organisms is of substantial interest. It is known that genomes of some organisms are of mixed ancestry. Some sequences that are viral in origin are now stably incorporated into the genome of eukaryotic organisms. The viral  
20 sequences themselves may have originated in another substantially unrelated species. An understanding of the ancestry of a gene can be important in drawing proper analogies between related genes and their translation products in other organisms.

There is a need for a better understanding of codon context  
25 effects on translation, and for a method for determining the appropriate codons for any desired translational effect. There is also a need for a method for identifying coding regions of the genome from nucleotide sequence data. There is also a need for a method for controlling protein folding and for insuring that a foreign gene will fold appropriately  
30 when expressed in a host. Genes altered or constructed in accordance with desired translational efficiencies would be of significant worth.

Another aspect of the practice of recombinant DNA techniques for the expression by microorganisms of proteins of industrial and pharmaceutical interest is the phenomenon of "codon

- 9 -

preference". While it was earlier noted that the existing machinery for gene expression in genetically transformed host cells will "operate" to construct a given desired product, levels of expression attained in a microorganism can be subject to wide variation, depending in part on specific alternative forms of the amino acid-specifying genetic code present in an inserted exogenous gene. A "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. That these forms provide the message for only 20 different amino acids (as well as transcription initiation and termination) means that some amino acids can be coded for by more than one codon. Indeed, some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon. For reasons not completely understood, alternative codons are not at all uniformly present in the endogenous DNA of differing types of cells and there appears to exist a variable natural hierarchy or "preference" for certain codons in certain types of cells.

As one example, the amino acid leucine is specified by any of six DNA codons including CTA, CTC, CTG, CTT, TTA, and TTG (which correspond, respectively, to the mRNA codons, CUA, CUC, CUG, CUU, UUA and UUG). Exhaustive analysis of genome codon frequencies for microorganisms has revealed endogenous DNA of E. coli most commonly contains the CTG leucine-specifying codon, while the DNA of yeasts and slime molds most commonly includes a TTA leucine-specifying codon. In view of this hierarchy, it is generally held that the likelihood of obtaining high levels of expression of a leucine-rich polypeptide by an E. coli host will depend to some extent on the frequency of codon use. For example, a gene rich in TTA codons will in all probability be poorly expressed in E. coli, whereas a CTG rich gene will probably highly express the polypeptide. Similarly, when yeast cells are the projected transformation host cells for expression of a leucine-rich polypeptide, a preferred codon for use in an inserted DNA would be TTA.

The implications of codon preference phenomena on recombinant DNA techniques are manifest, and the phenomenon may

- 10 -

serve to explain many prior failures to achieve high expression levels of exogenous genes in successfully transformed host organisms-a less "preferred" codon may be repeatedly present in the inserted gene and the host cell machinery for expression may not operate as efficiently.

- 5 This phenomenon suggests that synthetic genes which have been designed to include a projected host cell's preferred codons provide a preferred form of foreign genetic material for practice of recombinant DNA techniques.

10 Protein Trafficking

- The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to
- 15 predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their
- 20 biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence.

- Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct
- 25 destinations. Over the past few years the dissection of the molecular machinery for targeting and localization of proteins has been studied vigorously. Defined sequence motifs have been identified on proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins.

30

SUMMARY OF THE INVENTION

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid products, when introduced directly into muscle cells,

- 11 -

induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

#### BRIEF DESCRIPTION OF THE DRAWINGS

- 5                    Figure 1 shows the nucleotide sequence of the VIRa vector.  
                    Figure 2 is a diagram of the VIRa vector.  
                    Figure 3 is a diagram of the Vtpa vector.  
                    Figure 4 is the VUb vector  
                    Figure 5 shows an optimized sequence of the HCV core  
10    antigen.  
                    Figure 6 shows VIRa.HCV1CorePAb, Vtpa.HCV1CorePAb  
                    and VUb.HCV1CorePAb.  
                    Figure 7 shows the Hepatitis C Virus Core Antigen  
                    Sequence.  
15                    Figure 8 shows codon utilization in human protein-coding  
                    sequences (from Lathe et al.).  
                    Figure 9 shows an optimized sequence of the HCV E1  
                    protein.  
                    Figure 10 shows an optimized sequence of the HCV E2  
20    protein.  
                    Figure 11 shows an optimized sequence of the HCV E1 +E2  
                    proteins.  
                    Figure 12 shows an optimized sequence of the HCV NS5a  
                    protein.  
25                    Figure 13 shows an optimized sequence of the HCV NS5b  
                    protein.

#### DETAILED DESCRIPTION OF THE INVENTION

- 30                    This invention relates to novel formulations of nucleic acid  
                    pharmaceutical products, specifically nucleic acid vaccine products.  
                    The nucleic acid vaccine products, when introduced directly into muscle  
                    cells, induce the production of immune responses which specifically  
                    recognize Hepatitis C virus (HCV).

- 12 -

Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV),  
5 delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a  
10 new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH). Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of  
15 hepatocellular carcinoma. Thus, a need exists for an effective method for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of  
20 Flavivirus particles are known, and are discussed in M. A. Brinton, in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid  
25 bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever  
30 virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

- 13 -

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologs are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the generation of the encoded protein in situ in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

The standard techniques of molecular biology for preparing and purifying DNA constructs enable the preparation of the DNA therapeutics of this invention. While standard techniques of molecular biology are therefore sufficient for the production of the products of this invention, the specific constructs disclosed herein provide novel therapeutics which surprisingly produce cross-strain

- 14 -

protection, a result heretofore unattainable with standard inactivated whole virus or subunit protein vaccines.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1  $\mu$ g to 1 mg, and preferably about 10  $\mu$ g to 300  $\mu$ g is administered directly into muscle tissue. Subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, or inhalation delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided.

The DNA may be naked, that is, unassociated with any proteins, adjuvants or other agents which impact on the recipient's immune system. In this case, it is desirable for the DNA to be in a physiologically acceptable solution, such as, but not limited to, sterile saline or sterile buffered saline. Alternatively, the DNA may be associated with surfactants, liposomes, such as lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture, (see for example WO93/24640) or the DNA may be associated with an adjuvant known in the art to boost immune responses, such as a protein or other carrier. Agents which assist in the cellular uptake of DNA, such as, but not limited to, calcium ions, detergents, viral proteins and other transfection facilitating agents may also be used to advantage. These agents are generally referred to as transfection facilitating agents and as pharmaceutically acceptable carriers. As used herein, the term gene refers to a segment of nucleic acid which encodes a discrete polypeptide. The term pharmaceutical, and vaccine are used interchangeably to indicate compositions useful for inducing immune responses. The terms construct, and plasmid are used interchangeably. The term vector is used to indicate a DNA into which genes may be cloned for use according to the method of this invention.



The following examples are provided to further define the invention, without limiting the invention to the specifics of the examples.

V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment.

pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. We removed the entire *lac* operon from this vector, which was unnecessary for our purposes and may be detrimental to plasmid yields and heterologous gene expression, by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The ampicillin resistance marker was replaced with the neomycin resistance marker to yield vector V1Jneo.

- 16 -

An Sfi I site was added to VIJneo to facilitate integration studies. A commercially available 13 base pair Sfi I linker (New England BioLabs) was added at the Kpn I site within the BGH sequence of the vector. VIJneo was linearized with Kpn I, gel purified, blunted  
5 by T4 DNA polymerase, and ligated to the blunt Sfi I linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated VIJns. Expression of heterologous genes in VIJns (with Sfi I) was comparable to expression of the same genes in VIJneo (with Kpn I).

10 Vector VIRa (Sequence is shown in Figure 1; map is shown in Figure 2) was derived from vector VIR, a derivative of the VIJns vector. Multiple cloning sites (*Bgl*II, *Kpn*I, *Eco*RV, *Eco*RI, *Sal*I, and *Not*I) were introduced into VIR to create the VIRa vector to improve the convenience of subcloning. VIRa vector derivatives containing the  
15 tpa leader sequence and ubiquitin sequence were generated (Vtpa (Figure 3) and Vub (Figure 4), respectively). Expression of viral antigen from Vtpa vector will target the antigen protein into the exocytic pathway, thus producing a secretable form of the antigen proteins. These secreted proteins are likely to be captured by  
20 professional antigen presenting cells, such as macrophages and dendritic cells, and processed and presented by class II molecules to activate CD4<sup>+</sup> Th cells. They also are more likely to efficiently simulate antibody responses. Expression of viral antigen through VUb vector will produce a ubiquitin and antigen fusion protein. The uncleavable  
25 ubiquitin segment (glycine to alanine change at the cleavage site, Butt et al., JBC 263:16364, 1988) will target the viral antigen to ubiquitin-associated proteasomes for rapid degradation. The resulting peptide fragments will be transported into the ER for antigen presentation by class I molecules. This modification is attempted to enhance the class I  
30 molecule-restricted CTL responses against the viral antigen (Townsend et al, JEM 168:1211, 1988).

- 17 -

## EXAMPLE 2

### DESIGN AND CONSTRUCTION OF THE SYNTHETIC GENES

#### A. Design of Synthetic Gene Segments for HCV Gene Expression:

- 5                   Gene segments were converted to sequences having identical translated sequences (except where noted) but with alternative codon usage as defined by R. Lathe in a research article from *J. Molec. Biol.* Vol. 183, pp. 1-12 (1985) entitled "Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data: Theoretical and
- 10   Practical Considerations". The methodology described below was based on our hypothesis that the known inability to express a gene efficiently in mammalian cells is a consequence of the overall transcript composition. Thus, using alternative codons encoding the same protein sequence may remove the constraints on HCV gene expression.
- 15   Inspection of the codon usage within HCV genome revealed that a high percentage of codons were among those infrequently used by highly expressed human genes. The specific codon replacement method employed may be described as follows employing data from Lathe et al.:
- 20                   1.     Identify placement of codons for proper open reading frame.
2.     Compare wild type codon for observed frequency of use by human genes (refer to Table 3 in Lathe et al.).
3.     If codon is not the most commonly employed,
- 25   replace it with an optimal codon for high expression based on data in Table 5.
4.     Inspect the third nucleotide of the new codon and the first nucleotide of the adjacent codon immediately 3'- of the first. If a 5'-CG-3' pairing has been created by the new codon selection, replace it
- 30   with the choice indicated in Table 5.
5.     Repeat this procedure until the entire gene segment has been replaced.
6.     Inspect new gene sequence for undesired sequences generated by these codon replacements (e.g., "ATTTA" sequences,

- 18 -

inadvertent creation of intron splice recognition sites, unwanted restriction enzyme sites, etc.) and substitute codons that eliminate these sequences.

7. Assemble synthetic gene segments and test for improved expression.

#### B. HCV CORE ANTIGEN SEQUENCE

The consensus core sequence of HCV was adopted from a generalized core sequence reported by Bukh et al. (PNAS, 91:8239, 1994). This core sequence contains all the identified CTL epitopes in both human and mouse. The gene is composed of 573 nucleotides and encodes 191 amino acids. The predicted molecular weight is about 23 kDa.

The codon replacement was conducted to eliminate codons which may hinder the expression of the HCV core protein in transfected mammalian cells in order to maximize the translational efficiency of DNA vaccine. Twenty three point two percent (23.2%) of nucleotide sequence (133 out of 573 nucleotides) were altered, resulting in changes of 61.3% of the codons (117 out 191 codons) in the core antigen sequence. The optimized nucleotide sequence of HCV core is shown in Figure 5.

#### C. CONSTRUCTION OF THE SYNTHETIC CORE GENE

The optimized HCV core gene (Figure 5) was constructed as a synthetic gene annealed from multiple synthetic oligonucleotides. To facilitate the identification and evaluation of the synthetic gene expression in cell culture and its immunogenicity in mice, a CTL epitope derived from influenza virus nucleoprotein residues 366-374 and an antibody epitope sequence derived from SV40 T antigen residues 684-698 were tagged to the carboxyl terminal of the core sequence (Figure 6). For clinical use it may be desired to express the core sequence without the nucleoprotein 366-374 and SV40 T 684-698 sequences. For this reason, the sequence of the two epitopes is flanked by two *EcoRI* sites which will be used to excise this fragment of

- 19 -

sequence at a later time. Thus an embodiment of the invention for clinical use could consist of the VIRa.HCVI CorePAb, Vtpa.HCVI CorePAb, or VUb.HCVI CorePAb plasmids that had been cut with EcoRI, annealed, and ligated to yield plasmids

5 VIRa.HCVI Core, Vtpa.HCVI Core, and VUb.HCVI Core.

The synthetic gene was built as three separate segments in three vectors, nucleotides 1 to 80 in VIRa, nucleotides 80 to 347 (*Bst*XI site) in pUC18, and nucleotides 347 to 573 plus the two epitope sequence in pUC18. All the segments were verified by DNA  
10 sequencing, and joined together in VIRa vector.

#### D. HCV Gene Expression Constructs:

In each case, the junction sequences from the 5' promoter region (CMVintA) into the cloned gene is shown. The position at which  
15 the junction occurs is demarcated by a "/", which does not represent any discontinuity in the sequence.

The nomenclature for these constructs follows the convention: "Vector name-HCV strain-gene".

20

#### VIRa.HCVI.CorePAb

---IntA--AGA TCT ACC / ATG AGC--HCV.Core--GCC / GAA TTC GCT TCC--  
PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH---

25

#### Vtpa.HCVI.CorePAb

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH---

30

#### VUb.HCVI.CorePAb.

---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--  
HCV.Core--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH---

- 20 -

VIRa.HCV1.Core

---IntA--AGA TCT ACC / ATG AGC--HCV.Core--GCC / TAA A / GTC GAC--  
BGH---

5

Vtpa.HCV1.Core

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core--GCC / TAA A / GTC GAC--BGH---

VUb.HCV1.Core

10 ---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--  
HCV.Core--GCC / TAA A / GTC GAC--BGH---

E. OTHER SYNTHETIC HCV GENES

15 Using similar codon optimization techniques, synthetic  
genes encoding the HCV E1 (Figure 9), HCV E2 (Figure 10), HCV  
E1+E2 (Figure 11), HCV NS5a (Figure 12) and HCV NS5b (Figure 13)  
proteins were created.

## WHAT IS CLAIMED:

1. A synthetic polynucleotide comprising a DNA sequence encoding an HCV protein selected from the group consisting of HCV core protein, HCV E1 protein, HCV E1+E2 protein, HCV NS5a protein, HCV NS5b protein and fragments thereof, the DNA sequence comprising codons optimized for expression in a vertebrate host.
2. A plasmid vector comprising the polynucleotide of Claim 1, the plasmid vector being suitable for immunization of a vertebrate host.

3. The polynucleotide of Claim 1 which is HCV genotype I/1a core.

4. The polynucleotide of Claim 1 having the sequence

```

1 ATGAGCAGCA ACCGCAAGCT CCAAGAGAAg ACCAAgAGGA ACACCAAGG GAGGCGGAG GATCTAAAT TTCTCGGG 80
81 AGGCGACAT GTGGAAGGg TTACCTGCT GCGGAGGAGG GCGGCAAGG TGGGgGAGG ggcTACGgG AAGAGCTT 160
161 AGGCTTCCA gCGGAGGgG AGGAGGCAAG CCAATGCGAA GCGGAGGgT CCGGAGGgG GGTCTGCGT CAGGCTCGG 240
241 TACCGCTAGC GCGTATATG CAATTAAGGC TTGCGCTGG GCGGCTGCT gCTGTCGCTT AGGCTTGGG GCGGCGCTG 320
321 GCGGCGCAAG GACGCGCAGG GgAGCTTGGG gAAGCTGGG AAGCTTATG ACACGCTTAC GTGCGCTTT GCGGAGCTA 400
401 TGGCTTACAT GCGGCTGCT GgAGCTTGGG TgAGCTGGT ggcTACGCT CTGCGCTAT GGTTAAGCT gCTGAGCA 480
481 GCGCTTAAT ATGCTACTG CAAAGCTGGT GCGCTGCTT TGTGATTT CTTGCTGGG CTGCTCTCT GCTGAGCT 560
561 gCTGCTTCT GTC

```

5. The plasmid vector of Claim 2 having the sequence

```

1 GATATTGCT ATTGCGATT GATAGCTTG TATCATATG ATAATATTA CATTTATAT GGTCTATTC CAACATTAC 80
81 GGCATCTTA CATTAATTAT TGAATCTTA TTAATATTA TAAATTAAG GGTTCATTAT TATATAGCA TATATGAGT 160
161 TGGGCTTAC ATAACTTAC GTAAATGTC GCGCTGCTG CTACCTCAAC GAGCGCGCTT CATTAAGCTT AATAATTAAT 240
241 TATTTTCCA TAGTAAGGC AATAAGGACT TTGCATTAC CTCAATGCT GAGCTATTTA GGTATAACTT CCGCACTTGG 320
321 ATTACATCA GCTATATATA TGCCAACTAC GCGGCTTAT GACTTCATG AGGTAAATG GCGGCGCTG CATTATGCG 400
401 AATACATAC CTTATGAGC TTCTCTACTT GCGATACAT CTACTATTA GTCATGCTA TTACCATAT GATTCGCTTT 480
481 TGTACTACA TCAATGCGG TGTATAGCG TTGACTTAC GGGGATTTC AATTTTAC CCGATTGAG TAAATGAGT 560
561 TTTTCTTTC CACCAAAATG AAGCTGACTT TCAAAATAT GCTAACCACT GCGGCGCTT GAGCAAAATG GCGCTAGCT 640
641 GCTTACGTC GAGCTTAT ATAAAGAGG CCGCTTACT GAAGCTTAC ATGCGCTTA GAGCGATTC AGCTCTTTT 720
721 GAGCTGATA GAGACAGG GAGCTATTC AGCTTGGG GCGGCAAGG GTCATTGTA AGCGGATTG GCGCTTCAA 800
801 GATTAAGCTA AGTACGCTT ATAGACTTA TAGGCGACT GCTTGGCTT CTTATGATG CTATACTT TTTGCTTTC 880
881 GGTATATA CCGCGCTTC CTGAGTTAT AGTTATGCT ATAGCTTAC GTATAGCTT GGTATTATTA GCAATTATTA 960
961 CCACTCGCT ATTGCTAGC ATACTTTCA TTACTAATC ATAACATTA TTTTGGAG AACTTCTTT ATTGCTATA 1040
1041 TGGCAATCA CTTGCTTCA GAGACTTAC CCACTTCT ATTTTACAG GATGCGCTT CATTTATTA TTACAAATT 1120
1121 ACATATACA CAGACGCTT CCGAGTGGC GCGCTTTTA TTAAACATA GCGGCTATT CCACTGGAAT CTTGCTTAC 1200
1201 TTTTCTGAC ATGCTCTTT CTTGCTTAC GCGCTTCTT CTATATCCA GCGCTTCTT CATCTTCA CCGCTTAT 1280
1281 GTCCTGGC AGCTCTTTC TTCTAAGCT GCGCTTCA GCTAGTACA GCGGATTC CAGCAAGC AGCTTCTCT 1360
1361 ATAACTGCT GCGCTTACG TATCTCTTC AAAATAGCT TTTCTTATA AGACTTACG CTAATTTTC TTCTCTCT 1440
1441 AAGCTAGCG CAGAAACAG TGCAGGAGC TGACTTTTC GCTTCTTTC GCGCTTACG ACATAATTC TTAAGCTA 1520
1521 GTTACGCTG GAGCTAGCT TACTTTCAG ACTACTTTC GCGCTTCTG GCGCTTACG ACATAATTC CCGCTTAT 1600
1601 ACATCTTCT CTTTCTAT CTTTCTTTC CCACTTACG TCTTATGCT TGCATTACG ACATAATTC AGCTTCTAG 1680
1681 GAAATCTAC AGCAACAG ACAGTACG CCACTTCTT AACTTCTTC GCGCTTCA CATTTATTA CCGCTTAT 1760
1761 TCTTCTTAC GAGCTTCTG AGCTTCTTC TGAAGCTAC CAGTAACAG TTTTATAGT CCACTTACG CCGCTTAT 1840
1841 CCGCTTCTG TGTCTAGCT TGTCTCTCT ATGCTTCT TGTCTCTG CCGCTTCTT CAGCAAGCT AGCTTCTAG 1920
1921 AGCTTCTTC TGGCTGCTT GGTCTTTC CCGCTTCTT TGTCTCTG CCGCTTCTT CAGCAAGCT AGCTTCTAG 2000
2001 CCAAAATCT GCGCTTCTT ATTACATTC TGAAGCTTC CTTTCTTAC CCGCTTCTT AATATCTA CCGCTTCT 2080
2081 CTTTCTTAC GGTCTTCTG GCTTCTCT CATCTTCT TTTTCTTCA CAGCTTCTT TTTTCTTCT TCGCTTCT 2160
2161 GGTCTTCT TCTTCTTA CTTTCTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCT 2240
2241 ATGAGCTT GCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCT 2320
2321 CAGCTTCTT AGCTTCTTC ATGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT CCGCTTCTT 2400
2401 GAGCTTCTT CCGCTTCTT CCGCTTCTT TGTCTTAA

```

- 22 -

2481 GTATTCTGCG GGTTCGCTG GCGCAGCACA GCAAGCGGGA GATTCGCGAA GACAATAGCA GGCATGCTTG GCAATGCGCTG 2500  
 2561 GGTTCATGCG GTAGCGCGCG AGCGCGCTTA ATTAAAGCGG CAGCGCGCGT AGCGAGCTGC TGAAGAATTG AGCGCGCTTG 2520  
 2641 TGACATGCTA AAAAGCGCGG GTTCGCTGCG TTTTTCGATA GGTTCGCGCG GTTCAGCGAG CATTCACAAA ATTCAGCTTG 2540  
 2721 AACTCAGAGC TCGCGAAAGC GACAGCGACT ATAAAGATAC CAGCGCTTTC GCGCTGCAAG GTTCGCTGCG GGTTCGCTG 2560  
 2801 TTTCAGCGCT GCGCGTTAGC GATACGCTGT CCGCGCTTTC GGTTCGCGAA AGCGTGCATC TTTCGCAATG GTTCAGCTTG 2580  
 2881 AGTATGCTTA GTTCGCTGTA GGTTCGCTGC TCGAAGCTGC GGTTCGCGAA GCAAGCGCGG GTTCGCGCGG AGCGCTGCGT 2600  
 2961 GTTATTCGCT AACTATGCTG TCGATGCTAA CCGCGTAAGA CAGCGCTTAT GCAAGCTGCG AGCGAGCGCT GCTTAATAGCA 2620  
 3041 TTACAGAGAG GAGGTATGTA GCGCTGCTTA CAGAGCTTTC GAGCTGCTGC GCTAACTAGC GCTACGCTAG AAGGATAGTA 2640  
 3121 TTTGCTATGT GCGCTGCTGT GAACTGCTTT AGCTTCGCAA AAGAGCTTTC TAGCTGCTGA TCGCGCAAAA AAAAGCTGCT 2660  
 3201 TGCTAGCGCT GGTTCCTTTC TTTGCAAGCA GCGAGTTAGC GTACAGAAAA AAGGATGCTA AGAGATGCT TTGATGCTTT 2680  
 3281 CTAGCTGATC CGTAAAGCT GTTCGCTGCT TACAGCGAAT TAAGCAATTG TATTAGAAAA AACTGATGCA GATTAAGATT 2700  
 3361 AAAGTCAAT TTATTGATAT CAGGATTATC AATAGCGAAT TTTTCAAAAA GCGCTTTCG TAATTAAGCA GAAAACTGAT 2720  
 3441 GCGCGCGCTT GATAGCGATG GCAAGATGCT GGTATGCTG TCGCATGCG AGCTTCGCAA CATTAATACA AGTTATTGAT 2740  
 3521 TTGCGCTGCT CAAAAATAAG GTTATTAAGT GAAAAATGAC CATGCTGAG GATTAATGCG GGTGAGAAATG CCAAAAAGCTT 2760  
 3601 ATGATTTGCT TTTGAGCTT GTTAAAGCG CCAAGGATTA GGTTCGCTAT CAAAAATGCT GGTATTAAGT AAAGCTTTAT 2780  
 3681 TGATTGCTTA TTTGCGCTT GTTAAAGCA ATAGCTGATC GGTTCGAAAA GCAAAATTA\* AAAATGAAAT CCAATGCAAT 2800  
 3761 TTTGCGAGCA AGAGTGCAG GGTATTAACA ATATTGCTC GTGAAAGCG ATATTGCTT AATACTGCA ATGCTGTTTT 2820  
 3841 GCGCGGAT\* GCACTGCGA GTAAAGGATC ATGATGAGCA GTAGCGATAA AATGCTTAT GTTAAAGCA GGTATGAAAT 2840  
 3921 GGTTCAGCGA GTTTCGCTG AGCACTGCTC GTTAAAT\* GTTAAAGCG GTTTCGCTG GATGCTTTAG AAAGAACTG\* 2860  
 4001 GCGCGATGCG GGTTCGCTA CAATTGATAG ATTGCTGAG GTTATGCGCG GATATTATG GAGGCGGATA TATAGCGATA 2880  
 4081 TAAATAGCA TGTATGCTG AATTGATAG GGTTCGAG GTTATGCGCG GATATTATG ATGCTGATA AGAGCTTG 2900  
 4161 TATTACTGCT TATTAAAGCA GAGCTTTTA TGTTCATGA TATATATTT TTATTTGCG CAATTAAGA TACAGATTG 2920  
 4241 TGAGACAGAA GGTTCGCTG C

25 6. The polynucleotide of Claim 4 from which the PAb sequence has been removed.

30 7. The plasmid vector of Claim 5 from which the PAb sequence has been removed.

35 8. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 ng and 100 mg of the polynucleotide of Claim 1 into the tissue of the vertebrate.

9. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into the tissue of a vertebrate the polynucleotide of Claim 1.

40 10. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 1 and a pharmaceutically acceptable carrier.

45 11. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 1 into the tissue of said primate and concurrently administering interleukin-12 parenterally.



- 23 -

12. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation and effector functions including lymphokine secretion specific to HCV antigens which  
5 comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 1.

13. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of  
10 Claim 1 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

14. A pharmaceutical composition comprising the polynucleotide of Claim 1.  
15

15. A method of inducing an immune response comprising administering the polynucleotide of Claim 1 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant,  
20 recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

16. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1  
25 ng and 100 mg of the polynucleotide of Claim 2 into the tissue of the vertebrate.

17. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into  
30 the tissue of a vertebrate the polynucleotide of Claim 2.

18. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 2 and a pharmaceutically acceptable carrier.

- 24 -

19. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 2 into the tissue of said primate and concurrently administering  
5 interleukin 12 parenterally.

20. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation and effector functions including lymphokine secretion specific to HCV antigens which  
10 comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 2.

21. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of  
15 Claim 2 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

22. A pharmaceutical composition comprising the polynucleotide of Claim 2.  
20

23. A method of inducing an immune response comprising administering the polynucleotide of Claim 2 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant,  
25 recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

24. The vector of Claim 2 which is selected from V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb, VUb.HCV1CorePAb,  
30 V1Ra.HCV1Core, Vtpa.HCV1Core and VUb.HCV1Core.

25. A pharmaceutical composition comprising the vector of Claim 21.

- 25 -

26. The DNA sequence of Claim 1 selected from the group consisting of a nucleotide sequence shown in Figure 5, Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13.

1/22

1	GATATTGGCT	ATTGGCCATT	GCATACGTTG	TATCCATATC	ATAATATGTA	CATTTAAT	GGTCAATGTC	CAACATTACC	80
81	GCCATGTTGA	CATTGATTAT	TGACTAGTTA	TAAATAGTAA	TCAATTACGG	GGTCATTAGT	TCATAGCCCA	TATATGGAGT	160
161	TCCGCGTTAC	ATAACTTACG	GTAATGGCC	CGCCTGGCTG	ACCGCCCAAC	GACCCCGCC	CATTGACGTC	AATAATGACG	240
241	TATGTTCCCA	TAGTAACGCC	AATAGGGACT	TTCCATTGAC	GTCAATGGGT	GGAGTATTTA	CGGTAAACATG	CCCACCTGGC	320
321	AGTACATCAA	GTGTATCATA	TGCCAAGTAC	GCCCCCTATT	GACGTCAATG	ACGGTAAATG	GCCCCCGCTGG	CATTATGCCC	400
401	AGTACATGAC	CTTATGGGAC	TTTCCTACTT	GGCAGTACAT	CTACGTATTA	GTCAATGCTA	TTACCATGGT	GATGCGGTTT	480
481	TGGCAGTACA	TCAATGGGCG	TGGATAGCGG	TTTGACTCAC	GGGGATTTC	AAGTCTCCAC	CCCATTTGACG	TCAATGGGAG	560
561	TTTGTTTTGG	CACCAAAATC	AACGGGACTT	TCCAAAATGT	CGTAACAACT	CCGCCCCATT	GACGCAAAATG	GGCGGTAGGC	640
641	GTGTACGGTG	GGAGGCTAT	ATAAGCAGAG	CTCGTTTAGT	GAACCGTCAG	ATCGCCTGGA	GACGCCATCC	ACGCTGTTTT	720
721	GACCTCCATA	GAAGACACCG	GGACCGATCC	AGCCTCCGCG	GCCGGGAACG	GTGCATTGGA	ACGGGATTTC	CCCGTGCCAA	800
801	GAGTGACGTA	AGTACCGCCT	ATAGAGTCTA	TAGGCCACCC	CCCTTGGCTT	CTTATGCAIG	CTATACIGTT	TTTGGCTTGG	880
881	GGTCTATACA	CCCCGCTTC	CTCATGTTAT	AGGTGATGGT	ATAGCTTAGC	CTATAGGTGT	GGGTTATTGA	CCATTATTGA	960
961	CCACTCCCT	ATTGGTGACG	ATACITTTCA	TTACTAATCC	ATAACATGGC	TCCTTGGCCAC	AACICICTTT	ATTGGCTATA	1040
1041	TGCCAATACA	CTGTCTTTCA	GAGACTGACA	CGGACTCTGT	ATTTTACAG	GATGGGTCT	CATTATTAT	TTACAAATTC	1120
1121	ACATATACAA	CACCACCGTC	CCCAGTGCCC	GCAGTTTTTA	TTAAACATAA	CGTGGGATCT	CCACGCGAAT	CTCGGGTACG	1200
1201	TGTTCCGGAC	ATGGGCTCTT	CTCCGGTAGC	GGCGGAGCTT	CTACATCCGA	GCCCTGCTCC	CATGCCCTCCA	GCGACTCATG	1280
1281	GTGCTCGGC	AGCTCCTTGC	TCCTAACAGT	GGAGGCCAGA	CTTAGGCACA	GCACGATGCC	CACCACCACC	AGTGTGCCGC	1360
1361	ACAAGGCCGT	GGCGGTAGGG	TATGTGTCTG	AAAATGAGCT	CGGGGAGCGG	GCTTGCACCG	CTGACGCAAT	TGGAAGACTT	1440
1441	AAGGCAGCGG	CAGAAGAAGA	TGCAGGCAGC	TGAGTTGTTG	TGTTCTGATA	AGAGTCAGAG	GTAACCTCCG	TTGCGGTGCT	1520
1521	GTTAACGGTG	GAGGGCAGTG	TAGTCTGAGC	AGTACTCGTT	GCTGCCGCGC	GCGCCACCAG	ACATAATAGC	TGACAGACTA	1600
1601	ACAGACTGTT	CCTTTCCATG	GGCTTTTCT	GCAGTCACCG	TCCTTAGATC	TAGGTACCAG	ATATCAGAAT	TCAGTCGACA	1680
1680	GCGGCCGCGA	TCGTCTGTGC	CTTCTAGTTG	CCAGCCATCT	GTGTGTTGCC	CCTCCCCCGT	GCCTTCTTIG	ACCTTGAAG	1760
1761	GTGCCACTCC	CAGTGTCTTT	TCCTAATAAA	ATGAGGAAT	TGCATCGCAT	TGCTCTGAGTA	GGTGTCTTTC	TATTTCTGGG	1840
1841	GCTGGGGTGG	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAAATAGCAG	GCATGCTGGG	GATGCGGTGG	GCCTATGGG	1920
1921	TACGGCCGCA	GCGGCCTTAA	TTAAGGCCGC	AGCGGCCGTA	CCCAGGTGCT	GAAGAATTGA	CCCGGTTCTT	CGACCCGTAA	2000

FIG. 1A

2/22

2001	AAAGGCCGCG	TTGCTGGCGT	TTTTCCATAG	GCTCCGCCCC	CCTGACGAGC	ATCACAAAA	TCGACGCTCA	AGTCAGAGGT	2080
2081	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCCTG	2160
2161	CCGCTTACCG	GATACCTGTC	CGCCTTCTC	CCCTCGGGAA	GGTGCGGCT	TTCTCAATGC	TCAGGCTGTA	GGTATCTCAG	2240
2241	TTCGGTGTAG	GTCGTTGCT	CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	TTGAGCCCGA	CCGCTGCGCC	TTATCCGGTA	2320
2321	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	TAGCAGAGCG	2400
2401	AGGTATGTAG	GGGTGCTAC	AGAGTTCTTG	AAGTGTGGC	CTAACTACGG	CTACACTAGA	AGGACAGTAT	TTGGTATCTG	2480
2481	CGCTCTGCTG	AAGCCAGTTA	CCCTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	CCGGCAACAA	AACCACCGCT	GGTAGCGGTG	2560
2561	GTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TACGTGATCC	2640
2641	CGTAATGCTC	TGCCAGTGT	ACAACCAATT	AACCAATTCT	GATTAGAAAA	ACTCATCGAG	CATCAATGA	AACTGCAATT	2720
2721	TATTATATC	AGGATTATCA	ATACCATATT	TTTGAAAAAG	CCGTTTCTGT	AATGAAGGAG	AAACTCACC	GAGGCAGTTC	2800
2801	CATAGGATGG	CAAGATCCTG	GTATCGGICT	GCGATTCCGA	CTCGTCCAAC	ATCAATACAA	CCTATTAAAT	TCCCCTCGTC	2880
2881	AAAAATAAGG	TTATCAAGTG	AGAAATCACC	ATGAGTGACG	ACTGAATCCG	GTGAGAAATG	CAAAAGCTTA	TGCATTTCTT	2860
2961	TCCAGACTTG	TTCAACAGGC	CAGCCATTAC	GCTCGTCATC	AAAAATCCTC	GCATCAACCA	AACCGTTATT	CATTGCTGAT	3040
3041	TGCGCCTGAG	CGAGACGAAA	TACGCGATCG	CTGTAAAAAG	GACAAATTACA	AACAGGAAATC	GAATGCAACC	GGCGCAGGAA	3120
3121	CACTGCCAGC	GCATCAACAA	TATTTTCACC	TGAATCAGGA	TATCTTCTTA	ATACCTGGAA	TGCTGTTTTC	CCGGGGATCG	3200
3201	CAGTGGTGAG	TAACCATGCA	TCATCAGGAG	TACGGATAAA	ATGCTTGATG	GTGGAAGAG	GCATAAATTC	CGTCAGCCAG	3280
3281	TTTAGTCTGA	CCATCTCATC	TGTAACATCA	TTGGCAACGC	TACCTTTGCC	ATGTTTTCAGA	AACAACCTCTG	GCGCATCGGG	3360
3361	CTTCCCATAC	AATCGATAGA	TTGTCGCACC	TGATTGCCCG	ACATTATCGC	GAGCCCATTT	ATACCCATAT	AAATCAGCAT	3440
3441	CCATGTTGGA	ATTAAATCGC	GGCCTCGAGC	AAGACGTTTC	CCGTTGAATA	TGGCTCATAA	CACCCCTTGT	ATTACTGTTT	3520
3521	ATGTAAGCAG	ACAGTTTAT	TGTTTATGAT	GATAATTTT	TAICTTIGTC	AATGTAAACAT	CAGAGATTTT	GAGACACAAC	3600
3601	GTGGCTTTCC								3610

FIG.1B

3/22

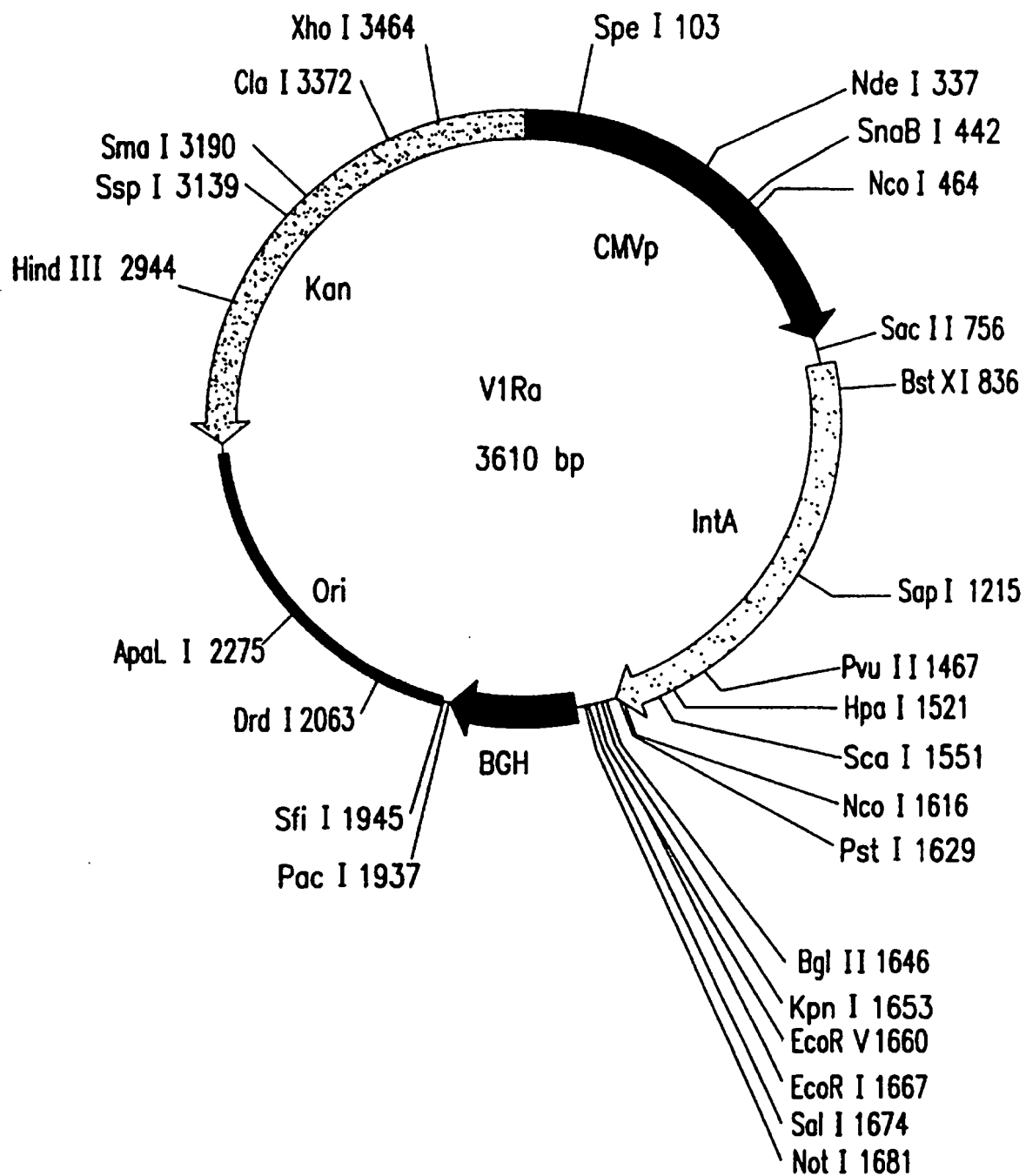


FIG.2

4/22

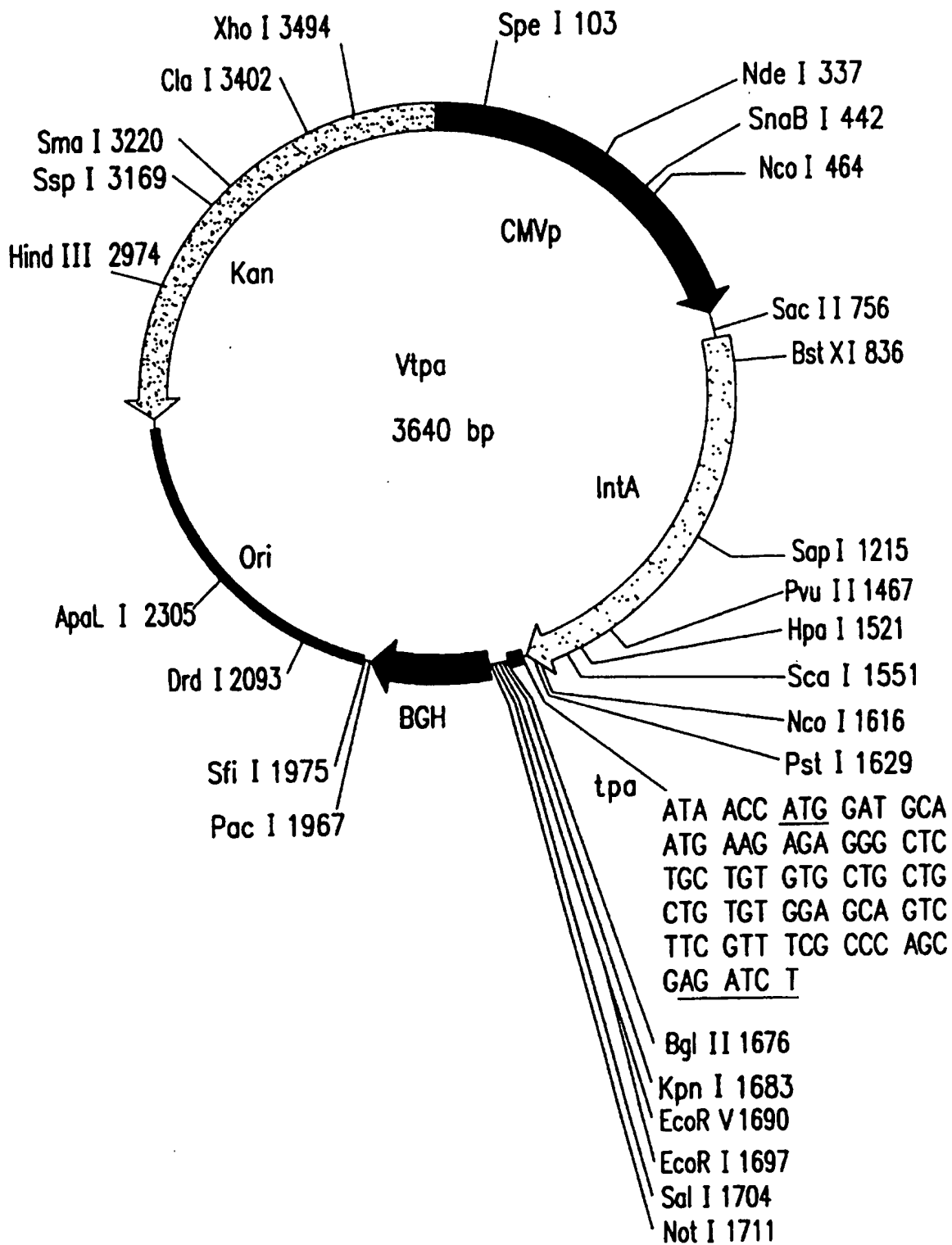


FIG.3

5/22

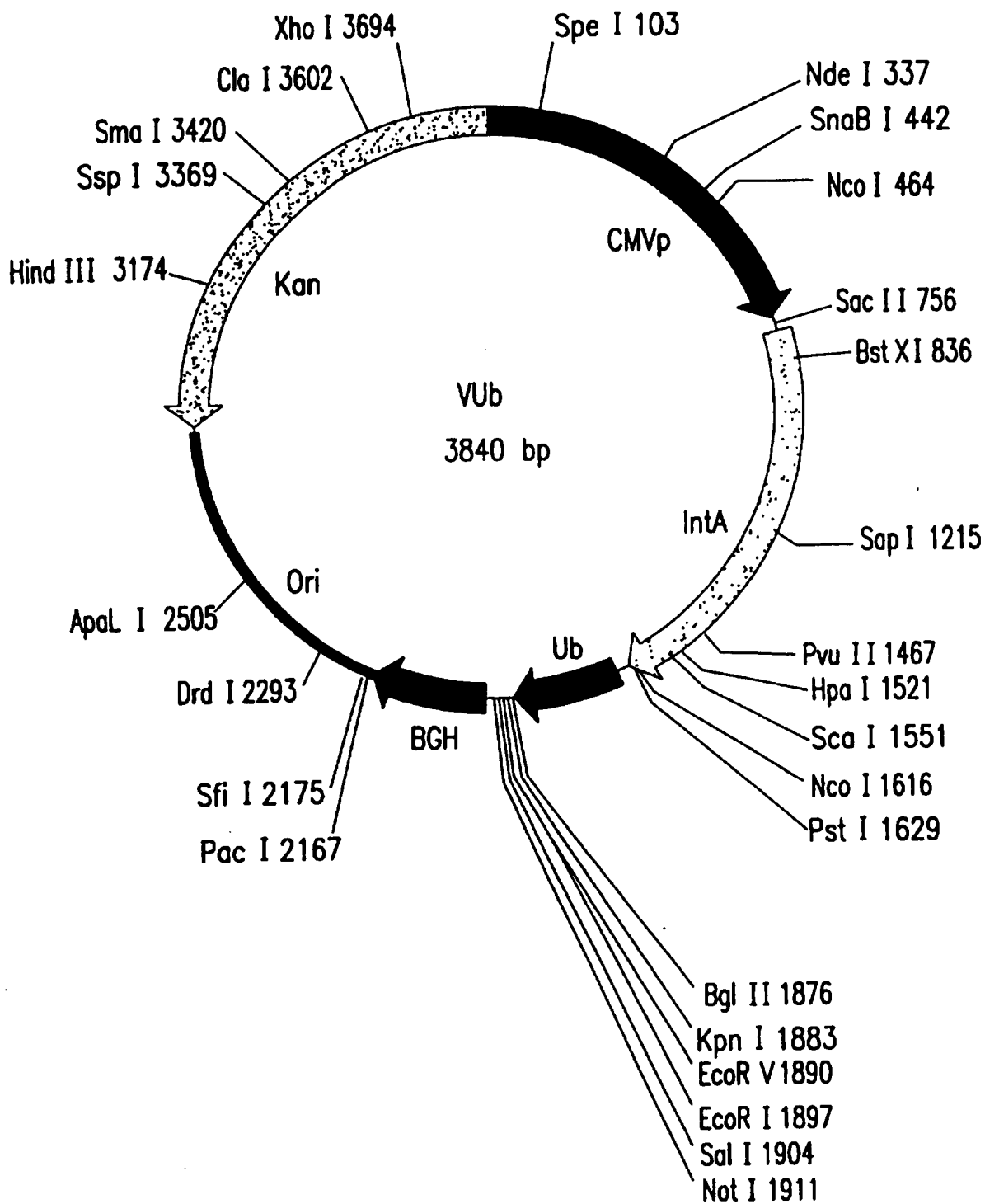


FIG.4



6/22

1/1 31/11  
 ATG AGC ACc AAC CCc AAg CCc CAg AGg AAg ACC AAg aGg AAC ACC AAC aGg aGg CCcCAG  
 Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr thr asn arg arg pro gln  
 61/21 91/31  
 GAT GTg AAG TTC CCT GGg GGa CAG ATt GTg GGa GGg GTc TAC cTg cTg CCc aGg AGG  
 asp val lys phe pro gly gly gln ile val gly gly val tyr leu leu pro arg arg  
 121/41 151/51  
 GGC CCC AGG cTg GGg GTg aGg GcT ACc aGg AAg ACC TcT GAG aGg TCc CAG CCC aGg GGC  
 gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly  
 181/61 211/71  
 AGG aGg CAG CCc ATC CCC AAG GCc aGg aGg CCT GAG GGC cGc TCc TGG GCc CAG CCT GGC  
 arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly  
 241/81 271/81  
 TAC CCc TGG CCC CTg TAT GGC AAT GAa GGC TtT GGC TGG GcT GGc TGG CTg CTg TCC CCC  
 try pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro  
 301/101 331/111  
 aGg GGC TCc aGg CCc tcc TGG GGC CCC ACa GAC CCC aGg aGg aGg TCc aGg AAC cTg GGC  
 arg gly ser arg pro ser trp gly pro thr asp pro arg arg ser arg asn leu gly  
 361/121 391/131  
 AAg GTg ATt GAc ACC CTg ACc Tgt GGC TtT GcT GAc CTg ATg GGc TAC ATC CCc CTg GTg  
 lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val  
 421/141 451/151  
 GGg GcT CCT GTg GGa GGg GTg GcT AGG GcT CTg GcT CAT GGg GTg AGG GTg CTg GAG GAT  
 gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp  
 481/161 511/171  
 GGG GTg AAC TAT GcT ACT GGC AAC cTg CCT GGC TGC TCc TTC ATC TTC CTg CTg GCC  
 gly val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala  
 541/181 571/191  
 CTg CTc TCC TGC CTg ACa GTg CCT GcT TCT GCc  
 leu leu ser cys leu thr val pro ala ser ala

FIG. 5

7/22

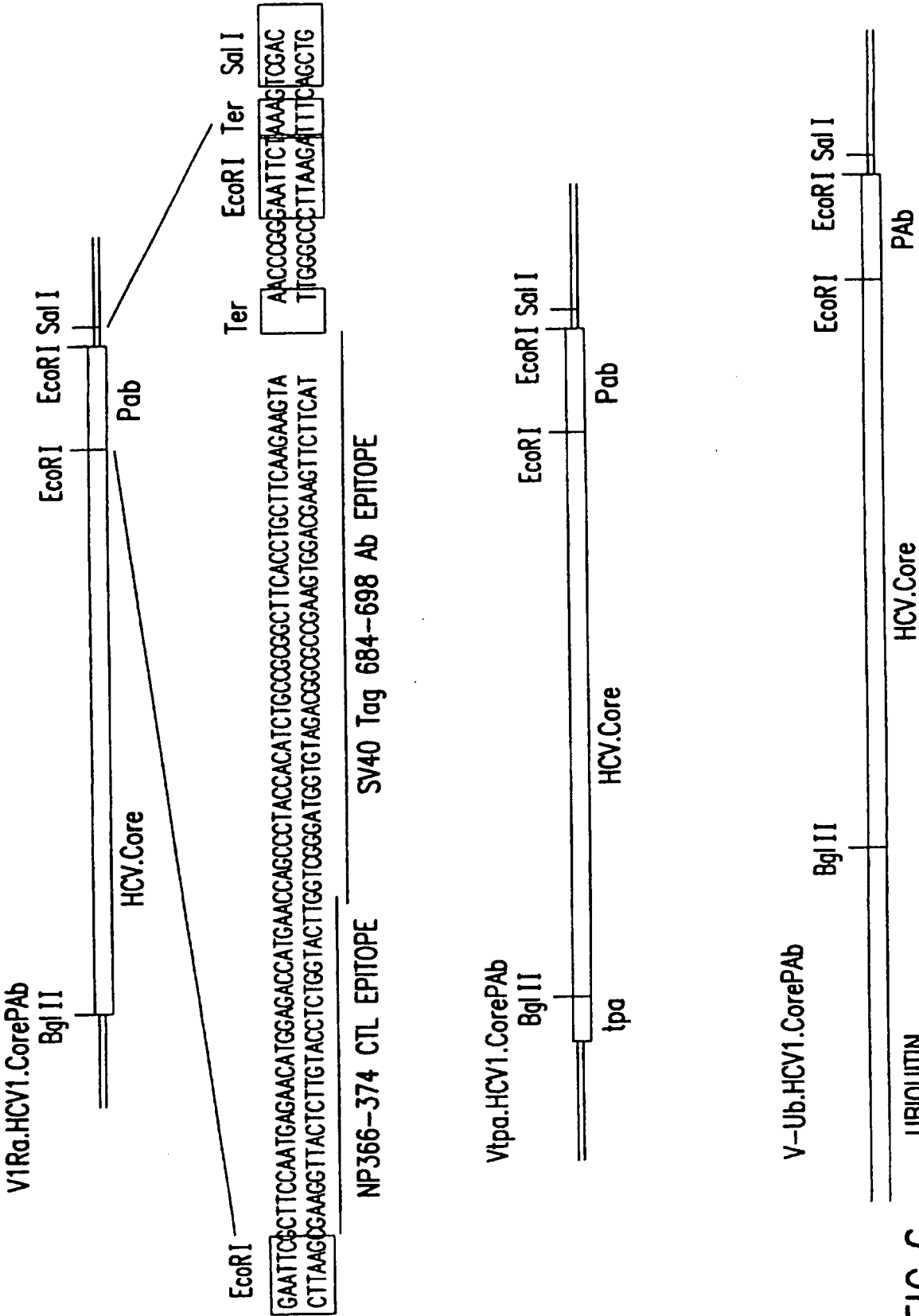


FIG. 6

1/1  
ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCa cAG  
Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr asn arg arg pro gln  
61/21 91/31  
GAC GTc AAG TTC CCg GGC Ggt GGT CAG ATC GTT GGA GTT TAC TTC TTG CCG CGC AGG  
asp val lys phe pro gly gly gln ile val gly gly val tyr leu leu pro arg arg  
121/41 151/51  
GGC CCC AGG TTG GGT GTG CGC GCG ACT aGG AAG ACT TcC GAG CGG TCG CAA CCT CGT GGA  
gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly  
181/61 211/71  
AGG CGa CAG CCT ATC CCC AAG Gct CGc CGG CCC GAG GGC AGG TCC TGG GCT CAG CCC GGG  
arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly  
241/81 271/91  
TAC Cct TGG CCC CTc Tat GGC AAT GAg GGC Ttc GGG TGG GCA GGA TGG CTC CTG TCC CCC  
tyr pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro  
301/101 331/111  
CGC GGC TCT CGg Cct agT TGG GGC CCc AcT GAC CCC CGG Cgt AGG TCG CGC AAT TTG GGT  
arg gly ser arg pro ser trp gly pro thr asp pro arg arg arg ser arg asn leu gly  
361/121 391/131  
AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC GCC GAC CTC ATG GgG TAC ATC CCG CTC GTC  
lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val  
421/141 451/151  
GGC GCC CCc GTA GgG GGC GTC GCC AGg GCC CTG CGC CAT GGC GTC AGG GtT cTG GAG GAC  
gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp  
481/161 511/171  
GGG gtg AAC TAT GCA ACA GGG Aat tIg cCc GGT TGC TCT TTC TCT ATC TTC CTC cTG Gct  
glu val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala  
541/181 571/191  
CTg CTg TcC TGC CTG ACC GTC CCA Gct TCT GCT  
leu leu ser cys leu thr val pro ala ser ala

**FIG. 7**

9/22

TABLE 3  
CODON UTILIZATION IN HUMAN PROTEIN-CODING SEQUENCES

a	b	c	d	e	f	a	b	c	d	e	f
F	UUU UUC	68 125	0.35 0.65	193	4.5	Y	UAU UAC	72 81	0.47 0.53	153	3.6
L	UUA UUG CUU CUC CUA CUG	20 42 50 99 30 204	0.05 0.09 0.11 0.22 0.07 0.46	445	10.4	H	CAU CAC	44 61	0.42 0.58	105	2.5
I	AUU AUC AUA	28 79 16	0.23 0.64 0.13	123	2.9	Q	CAA CAG	50 142	0.26 0.74	192	4.5
M	AUG	77	1.00	77	1.8	N	AAU AAC	51 97	0.34 0.66	148	3.5
V	GUU GUC GUA GUG	35 72 25 134	0.13 0.27 0.09 0.50	266	6.2	K	AAA AAG	137 166	0.45 0.55	303	7.0
S	UCU UCC UCA UCG AGU AGC	59 91 37 25 37 100	0.17 0.26 0.11 0.07 0.11 0.29	349	8.1	D	GAU GAC	79 130	0.38 0.62	209	4.9
P	CCU CCC CCA CCG	51 86 51 24	0.24 0.41 0.24 0.11	212	4.9	E	GAA GAG	125 186	0.40 0.60	311	7.3
T	ACU ACC ACA ACG	47 113 50 28	0.20 0.47 0.21 0.12	238	5.6	C	UGU UGC	44 103	0.30 0.70	147	3.4
A	GCU GCC GCA GCG	91 119 51 37	0.31 0.40 0.17 0.12	298	7.0	W	UGG	56	1.00	56	1.3
						R	CGU CGC CGA CGG AGA AGG	19 40 22 33 51 50	0.09 0.19 0.10 0.15 0.24 0.23	215	5.0
						G	GGU GGC GGA GGG	36 108 42 59	0.15 0.44 0.17 0.24	245	5.7

TOTAL 4285 RESIDUES EXCLUDING  
N-TERMINAL METHIONINE RESIDUES

FIG.8

10/22

1/1  
 atg TAT GAg GTG aGg AAt GTc Tct GGc GTc TAC CAT GTg Acc AAt GAC TGC TCC AAC TCC  
 M Y E V R N V S G V Y H V T N D C S N S  
 31/11  
 61/21  
 tGc ATT GTc TAT GAG Gct Gct GAC ATG ATc ATG CAC ACC CCT GGc Tgt GTg CCa Tgt GTg  
 C I V Y E A A D M I M H T P G C V P C V  
 91/31  
 121/41  
 aGg GAG GGc AAC TCC TCC aGg TGC TGG GTg GGc CTg ACC CCC Acc CTg Gct GCC AGG AAC  
 R E G N S S R C W V A L T P T L A A R N  
 151/71  
 181/61  
 tcC tcC ATC CCC Acc Acc ATc aGg aGg CAT GTg GAc cTG CTg GTg GGc Gct GCT GCC  
 S S I P T T I R R H V D L L V G A A A  
 211/71  
 241/81  
 CTg TGC Tct Gcc ATG TAT GTG GGc GAC CTg TGT GGc TCT GTc TTC CTg GTg TCC CAg gTG  
 L C S A M Y V G D L C G S V F L V S Q L  
 331/111  
 301/101  
 TTC ACC TTC TCC CCc aGg aGg TAT GAG Act GTg CAG GAC TGC AAC TGC TCC CTg TAC CCT  
 F T F S P R R Y E T V Q D C N C S L Y P  
 391/131  
 361/121  
 GGC CAT GTc Tct GGC CAC aGg ATG GCC TGG GAC ATG ATG ATG AAC TGG TCC CCc Acc Act  
 G H V S G H R M A W D M M N W S P T T  
 451/151  
 421/141  
 GCC cTg GTG GTc TCC CAG cTg CTg aGg ATt CCC CAG Gct GTg GTG GAC ATG GTG TGT Ggg  
 A L V V S Q L L R I P Q A V V D M V V G  
 511/171  
 481/161  
 GCC CAC TGG GGc GTg CTG Gct GGc CTg GCC TAC TAC TCC ATG GTG GGc AAC TGG GCC AAG  
 A H W G V L A G L A Y Y S M V G N W A K  
 571/191  
 541/181  
 GTg cTG ATT GTG ATG CTg CTg TTT Gct GGC GTg GAT GGc taa  
 V L I V M L L F A G V D G \*

FIG. 9

11/22

1/1  
 atg ACC ACC TAT GTc TcT GTG Ggc CAT Gcc tcC CAG ACC ACC aGg aGg GTg Gcc TCC TTC  
 M T Y V S V G H A S Q T T R R V A S F  
 31/11  
 61/21  
 TTC tcc CCT GGC TcT Gcc CAG AAg ATC CAG CTg GTg AAC ACC AAT GGC tcc TGG CAC ATC  
 F S P G S A Q K I Q L V N T N G S W H I  
 91/31  
 121/41  
 AAC AGG ACT GCC CTG AAT TGC AAT GAG TCC ATC AAC ATC GGC TTC TTT GcT Gcc CTG TTC  
 N R T A L N C N E S I N T G F F A A L F  
 151/51  
 181/61  
 TAT GTg AAG AAG TTC AAC TcC TCT GGC TGC TcT GAG aGg ATG Gcc tct Tgc aGg CCC ATT  
 Y V K K F N S S G C S E R M A S C R P I  
 211/71  
 241/81  
 GAC AGG TTt Gcc Cag Ggc TGG Ggc CCC ATC ACC CAT GCT GAG Tcc aGg tcc TcT GAC Cag  
 D R F A Q G W G P I T H A E S R S S D Q  
 271/91  
 301/101  
 AGG CcA TAC TGC TGG CAC TAT Gcc CcC CcC CcA TGT Ggc ATt GTG CcT Gcc cTG CAT GTc  
 R P Y C W H Y A P Q P C G I V P A L H V  
 331/111  
 361/121  
 Tgt Ggc CcT GTc TAc Tgc TTC ACC CcA tcC CCT GTg GTg Ggc ACg ACT GAC aGg TTt  
 C G P V Y C F T P S P V V G T T D R F  
 391/131  
 421/141  
 GGC GTg CCC ACc TAc AAC TGG Ggc GAC AAT GAG ACT GAT GTG CTg CTg AAC AAC ACC  
 G V P T Y N W G D N E T D V L L L N N T  
 451/151  
 481/161  
 aGg CCC CCc Cag Ggc AAC TGG TTt GGC Tgc Acc TGG ATG AAC tcC ACT Ggc TTC ACC AAG  
 R P P Q G N W F G C T W M N S T G F T K  
 511/171

FIG.10A

541/181	ACC Tgt GGC CCC Cca Tgc AAC Att Ggc Gct Ggc AAC ACC cTg ACC Tgc CCC
	T C G G P P C N I G A G N T L T C P
601/201	TTC aGG AAG Cat CCT GAG GCC Acc Tac ACC AAG Tgt GGC Tct GGc CCa TGG
	C F R K H P E A T Y T K C G S G P W
661/221	CCC AGG Tgc ATg GTg GAC Tac CCa Tac Agg CTg Tgg CAC Tac CCa Tgc Acc TTC
	L T P R C M V D Y P Y R L W H Y P C T F
721/241	ACC ATC TTc AAg ATC AGG ATg Tat TGt Ggc Ggc GTg GAG CAC AGG CTg Aat GCt
	N F T I F K I R M Y V G G V E H R L N A
781/261	AAC Tgc AAC Tgg Acc aGg Ggc GAg aGg Tgc AAC Att GAG GAc AGG TCt GAG CTg
	A C N W T R G E R C N I E D R D R S E L
841/281	tCc Ccc CTg CTg TCc Acc Act GAg Tgg Cag ATc CTg CCa Tgc TCC ACC ACC CTg
	S P L L L S T T E W Q I L P C S F T L
901/301	CTg TCC ACT GGc cTg ATC Cat CTg CAT CAG AAC Att GTg Gat GTg Cag Tac CTg
	P A L S T G L I H L H Q N I V D V Q Y L
961/321	Gtc Ggc Tct Gct GTg Gtc Tcc ATT GTg ATC Aag Tgg GAg Tat GTg CTg CTg
	Y G V G S A V V S I V I K W E Y V L L
1021/341	CTg CTg Gct Gat Gcc taa
	F L L L A D A *

FIG. 10B

13/22

1/1  
 atg TAT GAG GTG aGg AAt GTc Tct GGc GTc TAC CAT GTg Acc AAt GAC TGC TCC AAC TCc  
 M Y E V R N V S G V Y H V T N D C S N S  
 31/11  
 61/21  
 tGc ATT GTc TAT GAG Gct Gct GAC ATG ATc ATG CAC ACC Cct GGc Tgt GTg CCa Tgt GTg  
 C I V Y E A A D M I M H T P G C V P C V  
 91/31  
 121/41  
 aGG GAG GGc AAC TCC TCC aGg TGC TGG GTg GGc CTg Acc CCC Acc CTg Gct GCC AGG AAC  
 R E G N S S R C W V A L T P T L A A R N  
 151/51  
 181/61  
 tcC tcC ATC CCC Acc Acc ATc aGg aGg CAT GTg GAc cTG CTg GTg GGc Gct GCT GCC  
 S S I P T T T I R R H V D L L V G A A A  
 211/71  
 241/81  
 CTg TGC Tct GCC ATG TAT GTG GGc GAC CTg TGT GGc TCT GTc TTC CTg GTg TCC CAg cTG  
 L C S A M Y V G D L C G S V F L V S Q L  
 331/111  
 301/101  
 TTC ACC TTC TCC CCc aGg aGg TAT GAG ACT GTg CAG GAC TGC AAC TGC TCC CTg TAC CCt  
 F T F S P R R Y E T V Q D C N C S L Y P  
 391/131  
 361/121  
 GGC CAT GTc Tct GGc CAC aGg ATG GCC TGG GAC ATG ATG ATG AAC TGG TCC CCc Acc ACT  
 G H V S G H R M A W D M M N W S P T T  
 451/151  
 421/141  
 GCC cTg GTG GTc TCC CAG cTg CTg aGg ATt CCC CAG Gct GTg GTG GAC ATG GTG GTG GGc  
 A L V V S Q L L R I P Q A V V D M V V G  
 511/171  
 481/161  
 GCC CAC TGG GGc GTg CTG Gct GGc CTg GCC TAC TAC TCC ATG GTG GGc AAC TGG GCC AAG  
 A H W G V L A G L A Y Y S M V G N W A K

FIG.11A



14/22

541/181 571/191  
 GTg cTg ATT GTG ATG CTg CTg TTT Gct GGC GTg GAT GGC Acc ACC TAt GTc Tct GTG Ggc  
 V L I V M L L F A G V D G T T Y V S V G  
 601/201 631/211  
 CAT GCC tcC CAG ACC ACC aGg aGg GTg Gcc TCC TTC TTC tcc CCT GGC Tct GCC CAG AAg  
 H A S Q T T R R V A S F F S P G S A Q K  
 661/221 691/231  
 ATC CAg CTg GTg AAC ACC AAt GGC tcc TGG CAC ATC AAC AGG ACT GCC CTG AAt TGC AAt  
 I Q L V N T N G S W H I N R T A L N C N  
 721/241 751/251  
 GAG TCC ATC AAC ACT GGC TTC TTT Gct Gcc CTG TTC TAt GTg AAG AAG TTC AAC Tcc TCT  
 E S I N T G F F A A L F Y V K K F N S S  
 781/261 811/271  
 GGC TGC Tct GAG aGg ATG Gcc tct Tgc aGg CCC ATT GAC AGG TTt Gcc Cag GGC TGG GGC  
 G C S E R M A S C R P I D R F A Q G W G  
 841/281 871/291  
 CCC ATC ACC CAT GCT GAG Tcc aGg tcc Tct GAC CAg AGG CCa TAC TGC TGG CAC TAt Gcc  
 P I T H A E S R S S D Q R P Y C W H Y A  
 901/301 931/311  
 CCC CAg CCa TGT GGC ATt GTG Cct Gcc cTG CAT GTc Tgt Ggc Cct GTc TAc Tgc TTC ACC  
 P Q P C G I V P A L H V C G P V Y C F T  
 961/321 991/331  
 CCa tcC CCT GTg GTg GGC ACC Act GAC aGg TTt GGC GTg CCC ACc TAc AAC TGG GGC  
 P S P V V G T T D R F G V P T Y N W G  
 1021/341 1051/351  
 GAC AAt GAG Act GAt GTG CTg CTg AAC AAC ACC aGG CCC CcC Cag GGC AAC TGG TTe  
 D N E T D V L L L L N N T R P P Q G N W F

FIG. 11B

15/22

1081/361 1111/37  
 GGC Tgc Acc TGG ATg AAC tcc Act Ggc TTC ACC AAg Acc Tgt Ggc GGC CCC CCa Tgc AAC  
 G C T W M N S T G F T K T C G G P P C N  
 1141/381 1171/391  
 ATt Ggc Ggc Gct Ggc AAC AAC ACC cTg ACC Tgc CCC Act Gac Tgc TTC aGg AAg Cat Cct  
 I G G A G N N T L T C P T D C F R K H P  
 1201/401 1231/411  
 GAG GCC Acc TAC ACC AAg Tgt Ggc Tct Ggc CCa Tgg cTg Acc CCC AGG Tgc ATg GTg GAC  
 E A T Y T K C G S G P W L T P R C M V D  
 1261/421 1291/431  
 TAC CCa TAC Agg CTg TGG CAC TAC CCa Tgc Acc TTC AAC TTC ACC ATC TTC AAg ATC AGG  
 Y P Y R L W H Y P C T F N F T I F K I R  
 1321/441 1351/451  
 ATg TAT GTg Ggc Ggc GTg GAG CAC AGG CTg AAt Gct Gcc Tgc AAC Tgg Acc aGg Ggc GAG  
 M Y V G G V E H R L N A A C N W T R G E  
 1381/461 1411/471  
 aGg Tgc AAC ATg GAG CAC AGG GAC Agg Tct GAG CTg tcc Ccc CTg CTg CTg Tcc Acc Act  
 R C N I E D R D R S E L S P L L L S T T  
 1441/481 1471/491  
 GAG Tgg CAG ATc CTg CCa Tgc TCC TTC ACC ACC CTg Cct Gcc CTg TCC ACT Ggc cTg ATC  
 E W Q I L P C S F T T L P A L S T G L I  
 1501/501 1531/511  
 Cat CTg Cat CAG AAC ATt GTg GAT GTg Cag TAC CTg TAT Ggc GTg Ggc Tct Gct GTg GTc  
 H L H Q N I V D V Q Y L Y G V G S A V V  
 1561/521 1591/531  
 TCC ATT GTg ATC AAg TGG GAG TAT GTg CTg CTg TTC CTg CTg CTg Gct Gat Gcc taa  
 S I V I K W E Y V L L L F L L L A D A \*

FIG.11C

16/22

```

1/1      31/11
atg Tct Ggc Tcc Tgg CTg AGg GAT GTc TGG GAC TGG ATc TGC ACT GTG cTG ACT GAC TCC
M S G S W L R D V W D W I C T V L T D F

61/21    91/31
AAG ACC TGG CTg CAT TCC AAG CTg CTG Ccc aGG CTG Cct GGc GAC CCa TTC TTC Tcc TGc
K T W L H S K L L P R L P G D P F F S C

121/41   151/51
CAG aGg Ggc TAC AGG GGc GTc TGG aGG GGc GAT GGc GTg ATG CAG ACC ACC TGC CCa TGT
Q R G Y R G V W R G D G V M Q T T C P C

181/61   211/71
GGc Gcc CAG ATC ACT GGc CAT GTg AAg AAT GGc TCC ATG AGG ATt GTg GGc CCc AAg ACC
G A Q I T G H V K N G S M R I V G P K T

241/81   271/91
TGc tcc AAC ACC TGG CAT GGc ACC TTC CCC ATC AAt GCC TAC ACC ACT GGc CCa TGC ACC
C S N T W H G T F P I N A Y T T G P C T

301/101  331/111
CCa TCC Cct Gcc AAC TAC TCC AGG Gcc CTG TGG aG GTG GCT GCT GAG GAG TAT GTG
P S P A P N Y S R A L W R V A A E Y V

361/121  391/131
GAg GTg Acc aGG GTG GGc GAC TTC CAC TAt GTG ACT GGc ATG ACC ACT GAC AAt GTg AAg
E V T R V G D F H Y V T G M T T D N V K

421/141  451/151
TGC CCa TGC CAG GTg Cct GCC Cct GAg TTC TTC ACT GAg GTG GAT GGc GTG aGG cTG CAC
C P C Q V P P A P E F T E V D G V R L H

481/161  511/171
AGG TAt GCC Cct GCC TGC AAg CCc CTg CTg aGG GAT GAG GTg ACC TTC CAG GTg GGc CTg
R Y A P A C K P L L R D E V T F Q V G L

```

FIG. 12A

17/22

541/181 AAC CAG TTC Cct GTg Ggc Tcc CAG CTg CCa TGT GAG Cct GAg Cct GAT GTg Act GTG CTg  
 N Q F P V G S Q L P C E P E P D V T V L  
 571/191  
 601/201 ACC TCC ATG CTg Act GAg CCa TCC CAC ATc Act Gct GAG Act Gcc AAG aGg AGG cTG GCC  
 T S M L T T E P S H I T A E T A K R R L A  
 631/211  
 661/221 AGg Ggc Tcc Cct CCa TCC cTG GCC tcC Tcc TGcc tcC CAG cTG TCT Gct CCa Tcc cTG  
 R G S P P S L A S S S A S Q L S A P S L  
 691/231  
 721/241 AAG GCC ACC TGC ACC ACC aGg CAT GAC TCC Cct GAT Gct GAC CTg ATt GAG GCC AAC CTg  
 K A T C T T R H D S P D A D L I E A N L  
 751/251  
 781/261 CTG TGG aGg CAG GAG ATG GGC Ggc AAC ATc ACC aGg GTG GAG Tct GAG AAC AAG GTg GTg  
 L W R Q E M G G N I T R V E S E N K V V  
 811/271  
 841/281 ATc CTg GAC Tcc TTT GAg CCc CTg aGg Gct GAG GAG GAT GAG AGG GAg GTc Tct GTG Gct  
 I L D S F E P L R A E E D E R E V S V A  
 871/291  
 901/301 Gct GAG ATc CTg aGg AAg tcC AGG AAG TTC CCC Cct Gcc cTG CCC ATc TGG GcG aGg CCa  
 A E I L R K S R K F P P A L P I W A R P  
 931/311  
 961/321 tcc TAC AAC CCa CCc CTg CTg GAG Tcc TGG AAG GAC Cct GAC TAT GTg CCC Cct GTG GTg  
 S Y N P P L L E S W K D P D Y V P P V V  
 991/331  
 1021/381 CAT Ggc TGC CCc CTG CCc CCc ACC ATG GCC CCa CCC ATc CCc CCa CCC aGg AAG AGG  
 H G C P L P P T M A P P I P P P R R K R  
 1051/371

FIG.12B

18/22

1081/361  
 Act GTg GTg CTG Act GAg TCC Act GTc TCC TCT GCC cTG GCT GAG CTg GCC Acc AAG ACC  
 T V V L T E S T V S S A L A E L A T K T  
 1111/371  
 1141/381  
 TTC GGC tcC Tct GGC Tcc Tct Gct GTg GAC tct GGC Act Gcc ACG GCC CCC CCT GAC CAG  
 F G S S G S A V D S G T A T A P P D Q  
 1171/391  
 1201/401  
 CCa Tct GAT GAT GGC GAC AGg GGC Tct GAT GAT GAG Tcc TAC TCC TCC ATG CCC CCC CTg  
 P S D D G D R G S D D E S Y S S M P P L  
 1231/411  
 1261/421  
 GAG GGC GAG CCT GGC GAC Cct GAc CTg tct GAT GGC Tcc TGG Tcc Act GTc tct GAG GAG  
 E G E P G D P D L S D G S W S T V S E E  
 1291/431  
 1321/441  
 Gcc tct GAG GAT GTg GCC TGC TGC taa  
 A S E D V A C C S \*

FIG.12C

19/22

```

1/1      31/11
ATG TCc  TAC ACc TGG Act GGC GCC CTg ATC Acc CCa Tgt Gct Gct GAG GAG tCc AAG CTG
M  S  Y  T  W  T  G  A  L  I  T  P  C  A  A  E  E  S  K  L

61/21    91/31
CCC ATc  AAc CCc cTg tCc AAC TCc cTg CTg aGg Cat CAC AAC ATG GTc TAt GCC Acc Acc
P  I  N  P  L  S  N  S  L  L  R  H  H  N  M  V  Y  A  T  T

121/41   151/51
TCC aGg  tct Gct GGC CTg aGg CAG AAg AAG GTg ACC TTT GAC AGg CTG Cat GTg Cct Gat
S  R  S  A  G  L  R  Q  K  K  V  T  F  D  R  L  H  V  P  C

181/61   211/71
GAC CAC  TAC aGg Gat GTG CTg AAG GAG ATG AAG Gcc AAG Gcc TCC Act GTg AAG GCg AAg
D  H  Y  R  D  V  L  K  E  M  K  A  K  A  S  T  V  K  A  K

241/81   271/91
CTg CTg  TCT GTg GAg GGC TGC AAG CTG Acc Cct CCc CAC Tct GCC AGg TCC AAg TTT
L  L  S  V  E  E  A  C  K  L  T  P  P  H  S  A  R  S  K  F

301/101  331/111
GGC TAT  GGc AAC Gat GTg aGg AAC CTg TCC tCc AAG Gct GTg AAC CAC ATC CAC Tct
G  Y  G  A  K  D  V  R  N  L  S  S  K  A  V  N  H  I  H  S

361/121  391/131
GTc TGG  AAG GAC cTg CTg GAg GAC Act Gat Acc CCC ATT GAC Acc ACC ATC ATG GCc AAg
V  W  K  D  L  L  E  D  T  E  T  P  I  D  T  T  I  M  A  K

421/141  451/151
AAT GAg  GTc TTC TGT GTg CAg Cct GAG AAg GGc GGC aGg AAG Cct GCC aGg CTg AtT GTc
N  E  V  F  C  V  Q  P  E  K  G  G  R  K  P  A  R  L  I  V

481/161  511/171
TTC Cct  GAg CTg GGc GTg aGg GTG Tgt GAG AAg ATG GCC CTg TAt Gat GTG GTc TCC Acc
F  P  E  L  G  V  R  V  C  E  K  M  A  L  Y  D  V  V  S  T

```

FIG.13A

20/22

541/181 CAG Gct GTG ATG GGC TCC TCC TAT GGC TTC CAG TAC TCC CCT GGC CAG aGG GTg  
 CTg CCC L P Q A V M G S S Y G F Q Y S P G Q R V  
 571/191  
 601/201 CTG GTG AAT GCC TGG Aag TCC AAG Aag AAC CCC ATG GGC TTt GcC TAC TGC ACC  
 GAG TTC E F L V N A W K S K K N P M G F A Y C T  
 631/211  
 661/221 TTT GAC TCC Act GTg ACT GAG TcT GAc ATC aGg GTg GAG GAG TCC ATc TAC CAG  
 aGg TGC R C F D S T V T E S D I R V E E S I Y Q  
 691/231  
 721/241 GAC cTG Gct CCT GAG GCC AGg CAG GTg ATc AGG TCC CTg ACT GAG aGG CTg TAC  
 TGC TGT C C D L A P E A R Q V I R S L T E R L Y  
 751/251  
 781/261 GGC CCC CTG ACC AAC TCC Aag Ggc CAG AAC Tgt Ggc TAC aGg aGG TGC aGg GCC  
 ATt GGC I G G P L T N S K G Q N C G Y R R C R A  
 811/271  
 841/281 GTG CTG ACC ACT AAC Tgt Ggc AAC ACC CTg ACC Tgc TAC cTG AAG GCC TCT GCT  
 tct Ggc S G V L T T N C G N T L T C Y L K A S A  
 871/291  
 901/301 GCT GcC AAG CTg CAT GAC TGC ACC ATG CTg GTc Tgt Ggc Gat GAC CTg GTg  
 Gct TGc A C R A A K L H D C T M L V C G D D L V  
 931/311  
 961/321 TGT GAg tct Gct Ggc ACC CAG GAG Gat Gct GcC tcC CTg aGg GTc TTC Act GAG  
 GTg ATC V I C E S A G T Q E D A A S L R V F T E  
 991/331  
 1021/341 ACC AGG TAC TCT GCC CCC CCT GGc GAC CCT CCC CAG CCT GAg TAT GAC cTG GAG  
 GCC ATG A M T R Y S A P P G D P P Q P E Y D L E  
 1051/351

FIG.13B

21/22

1081/361 1111/371  
 cTg ATc Acc Tcc Tgc Tcc Tcc Aat GTc Tct GTg Gcc Cat Gat Gcc Tct Ggc Aag aGg GTc  
 L I T S C S S N V S V A H D A S G K R V  
 1141/381 1171/391  
 TAc TAc CTg Acc aGg GAc Ccc Acc CCc CTg Gcc AGg Gct Gcc TGG GAg Act Gcc AGg  
 Y Y L T R D P T T P L A R A A W E T A R  
 1201/401 1231/411  
 CAC Acc Cct GTg AAC Tcc Tgg CTg Ggc AAC ATc ATc ATg TAt Gcc Ccc Acc CTg TGG Gcc  
 H T P V N S W L G N I I M Y A P T L W A  
 1261/421 1291/431  
 AGG ATg ATc CTg ATg Acc CAC TTC TTC Tcc ATc CTg CTg Gcc CAG GAg CAg CTg GAg AAg  
 R M I L M T H F F S I L L A Q E Q L E K  
 1321/441 1351/451  
 GCC CTg Ggc Tgc CAG ATt TAt Ggc GCC Acc TAc TTC ATT GAg CCc CTg GAc CTg CCc CAG  
 A L G C Q I Y G A T Y F I E P L D L P Q  
 1381/461 1411/471  
 ATc ATc CAG aGg CTg Cat Ggc CTg tct Gcc Ttc Tcc CTg CAC tcc TAc Tcc Cct Ggc GAg  
 I I Q R L H G L S A F S L H S Y S P G E  
 1441/481 1471/491  
 ATc AAC AGg GTg Gcc Tcc Tgc CTg AGg AAg CTg Ggc GTg Ccc Ccc cTg aGg GTg TGG AGg  
 I N R V A S C L R K L G V P P L R V W R  
 1501/501 1531/511  
 GAc aGg GCC AGg tct GTg aGg GCC AAg CTg CTg Tcc CAG Ggc Ggc AGg Gct GCC Acc TGT  
 H R A R S V R A K L L S Q G G R A A T C  
 1561/521 1591/531  
 GGC AAg TAc CTg TTC AAC TGG Gct GTg AGg Acc AAg CTg AAg CTg Acc Ccc ATc Cct GCT  
 G K Y L F N W A V R T K L K L T P I P A

FIG.13C



22/22

```

1621/541      1651/551
GCc TCC CAG cTg GAC cTg Tct GGT GTg GGT TAC tct GGC GGC GAC ATc TAC
A S Q L D L S G W F V A G Y S G G D I Y
1681/561      1711/571
CAC tcC CTG Tcc aGg GCC aGg CCC aGg TGG TTC ATG TGG TGC CTg CTg CTg TCT GTg
H S L S R A R P R W F M W C L L L L S V
1741      1771/591
GGc GTg GGC ATC TAC CTG CTg CCC AAC aGG TGA
G V G I Y L L P N R *
```

FIG.13D

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : A61N 43/04; C12Q 1/68; C12N 15/00; C07H 21/02; A61K 39/00

US CL : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, MEDLINE, BIOSIS, EMBASE, CAPLUS, WPIDS, SCISEARCH

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Selby et al. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. Journal of General Virology. 1993. Vol. 74, pages 1103-1113, see entire document.	1-3
X	Bukh et al. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc. Natl. Acad. Sci. August 1994. Vol. 91, pages 8239-8243, see entire document.	1-3
Y	Lathe. Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data Theoretical and Practical Considerations. J. Mol. Biol. 1985. Vol. 183, pages 1-12, see entire document.	1-3



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

28 AUGUST 1997

Date of mailing of the international search report

11 SEP 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ANDREW WANG

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Grantham et al. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Research. 1981. Vol. 9, No. 1, pages r43-r74, see entire document.	1-3
A, P	Ide et al. Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. Gene. December 1996. Vol. 182, pages 203-211, see entire document.	1-3, 8-26
X	US 5,514,539 A (BUKH et al.) 07 May 1996, see entire document.	1-3, 8-26

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☒ Claims Nos.: 4-7  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
  
The enclosed copy of claims 4, 5 were not legible and claims 6, 7 depend on those claims.
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

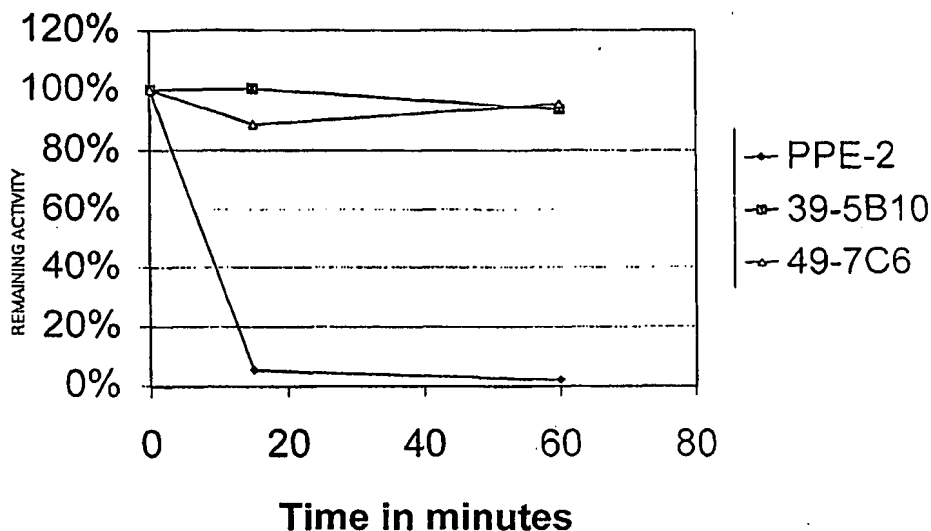


## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : C12N 15/53, 9/02, C12Q 1/68		A2	(11) International Publication Number: <b>WO 99/14336</b>
			(43) International Publication Date: 25 March 1999 (25.03.99)
(21) International Application Number: PCT/US98/19494		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 18 September 1998 (18.09.98)			
(30) Priority Data: 60/059,379 19 September 1997 (19.09.97) US			
(71) Applicant (for all designated States except US): PROMEGA CORPORATION [US/US]; 2800 Woods Hollow Road, Madison, WI 53711-5399 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): WOOD, Keith, V. [US/US]; 902 Kottke Drive #5, Madison, WI 53719 (US). HALL, Mary, P. [US/US]; 6305 Dylun Drive, Madison, WI 53719 (US).		Published Without international search report and to be republished upon receipt of that report.	
(74) Agent: MARTIN, Alice, O.; Brinks Hofer Gilson & Lione, P.O. Box 10087, Chicago, IL 60610 (US).			

(54) Title: THERMOSTABLE LUCIFERASES AND METHODS OF PRODUCTION

## Stability at 37C normalized to t=0



## (57) Abstract

Luciferase enzymes with greatly increased thermostability, e.g., at least half lives of 2 hours at 50 °C, cDNAs encoding the novel luciferases, and hosts transformed to express the luciferases, are disclosed. Methods of producing the luciferases include recursive mutagenesis. The luciferases are used in conventional methods, some employing kits.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## THERMOSTABLE LUCIFERASES AND METHODS OF PRODUCTION

5           The government may have rights to this invention based on support provided by NIH 1R43 GM506 23-01 and 2R44 GM506 23-02 and NSF ISI-9160613 and III-9301865.

## RELATED APPLICATIONS

10           This application claims priority from copending U.S. Ser. No. 60/059,379 filed September 19, 1997.

## FIELD OF THE INVENTION

15           The invention is directed to mutant luciferase enzymes having greatly increased thermostability compared to natural luciferases or to luciferases from which they are derived as measured e.g. by half-lives of at least 2 hrs. at 50°C in aqueous solution. The invention is also drawn to polynucleotides encoding the novel luciferases, and to hosts transformed to express the luciferases. The invention is further drawn to methods of producing luciferases with increased thermostability and the use of these luciferases in any method in which previously known luciferases are conventionally employed. Some of the uses employ kits.

## 20           BACKGROUND OF THE INVENTION

          Luciferases are defined by their ability to produce luminescence. Beetle luciferases form a distinct class with unique evolutionary origins and chemical mechanisms. (Wood, 1995)

25           Although the enzymes known as beetle luciferases are widely recognized for their use in highly sensitive luminescent assays, their general utility has been limited due to low thermostability. Beetle luciferases having amino acid sequences encoded by cDNA sequences cloned from luminous beetles are not stable even at moderate temperatures. For example, even the most stable of the luciferases, *LucPpe2*, obtained from a firefly has very little stability at the

moderate temperature of 37° C. Firefly luciferases are a sub-group of the beetle luciferases. Historically, the term "firefly luciferase" referred to the enzyme LucPpy from a single species *Photinus pyralis* (Luc + is a version).

Attempts have been reported to mutate natural cDNA sequences encoding luciferase and to select mutants for improved thermostability (White et al., 1994; from *P. pyralis* and Kajiyama and Nekano, 1993, from *Luciola lateralis*.) However, there is still a need to improve the characteristics and versatility of this important class of enzymes.

## SUMMARY OF THE INVENTION

The invention is drawn to novel and remarkably thermostable luciferases, including half-lives of at least 2 hrs. at 50°C or at last 5 hrs. at 50°C in aqueous solution. The mutant luciferases of the present invention display remarkable and heretofore unrealized thermostability at room temperature (22°C) and at temperatures at least as high as 65°C. The invention is further directed to the mutant luciferase genes (cDNA) which encode the novel luciferase enzymes. The terminology used herein is, e.g. for the mutants isolated in experiment 90, plate number 1, well B5, the *E. coli* strain is 90-1B5, the mutant gene is *luc90-1B5*, and the mutated luciferase is Luc90-1B5.

By thermostability is meant herein the rate of loss of enzyme activity measured at half life for an enzyme in solution at a stated temperature. Preferably, for beetle luciferases, enzyme activity means luminescence measured at room temperature under conditions of saturation with luciferin and ATP. Thermostability is defined in terms of the half-life (the time over which 50% of the activity is lost).

The invention further encompasses expression vectors and other genetic constructs containing the mutant luciferases, as well as hosts, bacterial and otherwise, transformed to express the mutant luciferases. The invention is also drawn to compositions and kits which contain the novel luciferases, and use of these luciferases in any methodology where luciferases are conventionally employed.



Various means of random mutagenesis were applied to a luciferase gene (nucleotide sequence), most particularly gene synthesis using an error-prone polymerase, to create libraries of modified luciferase genes. This library was expressed in colonies of *E. coli* and visually screened for efficient luminescence to select a subset library of modified luciferases. Lysates of these *E. coli* strains were then made, and quantitatively measured for luciferase activity and stability. From this, a smaller subset of modified luciferases was chosen, and the selected mutations were combined to make composite modified luciferases. New libraries were made from the composite modified luciferases by random mutagenesis and the process was repeated. The luciferases with the best overall performance were selected after several cycles of this process.

Methods of producing improved luciferases include directed evolution using a polynucleotide sequence encoding a first beetle luciferase as a starting (parent) sequence, to produce a polynucleotide sequence encoding a second luciferase with increased thermostability, compared to the first luciferase, while maintaining other characteristics of the enzymes. A cDNA designated *lucppe2* encodes a firefly luciferase derived from *Photuris pennsylvanica* that displays increased thermostability as compared to the widely utilized luciferase designated *LucPpy* from *Photinus pyralis*. The cDNA encoding *LucPpe2* luciferase was isolated, sequenced and cloned (see Leach, *et al.*, 1997). A mutant of this gene encodes a first luciferase *LucPpe2* [T249M].

In an embodiment of a mutant luciferase, the amino acid sequence is that of *LucPpe2* shown in FIG. 45 with the exception that at residue 249 there is a T (designated T249 M) rather than the M reported by Leach *et al.* The bold, underlined residue (249) shows mutation from T to M. This enzyme produced approximately 5-fold more light *in vivo* when expressed in *E. coli*. Double-underlined residues were randomized by oligonucleotide mutagenesis.

Diluted extracts of recombinant *E. coli* that expressed mutant luciferases made by the methods of the invention were simultaneously screened for a plurality of characteristics including light intensity, signal stability, substrate utilization ( $K_m$ ), and thermostability. A fully automated robotic system was used to screen

large numbers of mutants in each generation of the evolution. After several cycles of mutagenesis and screening, thereby creating mutant libraries of luciferases, an increased thermostability compared to *LucPpe2* [T249M] of about 35°C was achieved for the most stable clone [clone *Luc90-1B5*] which also essentially maintained thermostability (there was only negligible loss in activity of 5%) when kept in aqueous solution over 2 hrs. at 50°C, 5 hours at 65°C, or over 6 weeks at 22°C.

Mutant luciferases of the present invention display increased thermostability for at least 2 hrs. at 50°C, preferably at least 5 hrs. at 50°C in the range of 2-24 hrs. at 50°-65°C. In particular, the present invention comprises thermostable mutant luciferases which, when solubilized in a suitable aqueous solution, have a stability half-life greater than about 2 hours at about 50°C, more preferably greater than about 10 hours at 50°C, and more preferably still greater than 5 hours at 50°C. The present invention also comprises mutant luciferases which, when solubilized in a suitable aqueous solution, have a stability half-life greater than about 5 hours at about 60°C, more preferably greater than about 10 hours at about 60°C, and more preferably still greater than about 24 hours at about 60°C. The present invention further comprises mutant luciferases which when solubilized in a suitable aqueous solution have a stability half-life greater than about 3 months at about 22°C, and more preferably a half-life stability of at least 6 months at 22°C. An embodiment of the invention is a luciferase mutant having stability 6 hours at 65°C (equivalent to a half-life of 2 days). A loss of activity of about 5-6% was found. The half-lives of enzymes from the most stable clones of the present invention, extrapolated from data showing small relative changes, is 2 days at 65°C (corresponding to 6% loss over 6 hours), and 2 years at 22°C (corresponding to 5% loss over 6 weeks).

In particular, the invention comprises luciferase enzymes with embodiments of amino acid sequences disclosed herein, (e.g. mutant luciferases designated *Luc49-7C6*; *Luc78-0B10*; and *Luc90-1B5*, FIGS. 27, 36, 43) as well as all other beetle luciferases that have thermostability as measured in half-lives of at

least 2 hours at 50°C. The invention also comprises mutated polynucleotide sequences encoding luciferase enzymes containing any single mutation or any combination of mutations of the type and positions in a consensus region of beetle luciferase encoding sequences, disclosed herein, or the equivalents. The mutations are indicated in the sequences in FIGS. 22-47 by bold, underlined residues and are aligned with other beetle luciferase sequences in FIG. 19.

Nucleotide sequences encoding beetle luciferases are aligned in FIG. 19. Eleven sequences found in nature in various genera and species within genera are aligned, including *lucppe-2*. Nucleotide sequences encoding three mutant luciferases of the present invention (*Luc49-7C6*; *78-0B10*; *90-1B5*) are also aligned. There are at least three mutations in each mutant luciferase that show increased thermostability. In general, mutations are not in the conserved regions. Conserved amino acids are those that are identical in all natural species at positions shown in FIG. 19. Consensus refers to the same amino acid occurring at more than 50% of the sequences shown in FIG. 19, excluding *LucPpe2*.

#### DETAILED DESCRIPTION OF THE INVENTION

The invention relates beetle luciferases that are characterized by high thermostability and are created by mutations made in the encoding genes, generally by recursive mutagenesis. The improved thermostability allows storage of luciferases without altering its activity, and improves reproducibility and accuracy of assays using the new luciferases. The invention further comprises isolated polynucleotide sequences (cDNAs) which encode the mutant luciferases with increased thermostability, vectors containing the polynucleotide sequences, and hosts transformed to express the polynucleotide sequences. Table 1 shows results of about 250 clones and characteristics of the luciferases from the clones including thermostability. The invention also encompasses the use of the mutant luciferases in any application where luciferases are conventionally utilized, and kits useful for some of the applications.

Unexpectedly, beetle luciferases with the sought after high thermostability were achieved in the present invention through a process of recursive mutagenesis

and selection (sometimes referred to as "directed evolution"). A strategy of recursive mutagenesis and selection is an aspect of the present invention, in particular the use of a multi-parameter automated screens. Thus, instead of screening for only a single attribute such as thermostability, simultaneous  
5 screening was done for additional characteristics of enzyme activity and efficiency. By this method, one property is less likely to "evolve" at the expense of another, resulting in increased thermostability, but decreased activity, for example.

Table 1 presents examples of parameter values ( $L_i$ ,  $\tau$ ,  $K_m$  and  $S$ ) derived  
10 from experiments using different luciferases as starting (parent) sequences. The subtitles refer to designations of the starting temperature at which the parameters were measured and the starting luciferase, e.g., "39-5B10 at 51°C" and so forth. All parameters in each experiment are recorded as relative values to the respective starting sequence, e.g., the parameter values for the starting sequence in any  
15 experiment equal "1." (See Example 2 herein for definitions.)

Thermostability has evolved in nature for various enzymes, as evidenced by thermostable isozymes found in thermophilic bacteria. Natural evolution works by a process of random mutagenesis (base substitutions, gene deletions, gene insertions), followed by selection of those mutants with improved  
20 characteristics. The process is recursive over time. Although the existence of thermostable enzymes in nature suggests that thermostability can be achieved through mutagenesis on an evolutionary scale, the feasibility of achieving a given level of thermostability for a particular class of enzymes by using short term laboratory methods was unpredictable. The natural process of evolution, which  
25 generally involves extremely large populations and many millions of generations and genes, by mutation and selection cannot be used to predict the capabilities of a modern laboratory to produce improved genes by directed evolution until such mutants are produced.

After such success, since the overall three-dimensional structure of all  
30 beetle luciferases are quite similar, having shown it possible for one member of this class makes it predictable that high thermostability can be achieved for other

beetle luciferases by similar methods. FIG. 17 shows evolutionary relationship among beetles luciferases. All of these have a similar overall architecture. The structural class to which the beetle luciferases belong is determined by the secondary structure (e.g. helices are symbolized by cylinders, sheets by collections of arrows, loops connect helices with sheets (FIG. 18A). FIG. 18B shows the amino acids of the *LucPpe2* luciferase (FIG. 18B) wherein small spirals correspond to cylinders of FIG. 18A; FIG 18C shows that the general beetle architecture matches (is superimposed on) that of *LucPpe2*. This is support for the expectation that the methods of the present invention may be generalized to all beetles luciferases:

Enzymes belong to different structural classes based on the three-dimensional arrangement of secondary elements such as helices, sheets, and loops. Thermostability is determined by how efficiently the secondary elements are packed together into a three-dimensional structure. For each structural class, there also exists a theoretical limit for thermostability. All beetle luciferases belong to a common structural class as evident by their common ancestry (FIG. 17), homologous amino acid sequences, and common catalytic mechanisms.

The application of a limited number of amino acid substitutions by mutagenesis is unlikely to significantly affect the overall three-dimensional architecture (*i.e.*, the structural class for mutant luciferases is not expected to change.) Because the theoretical limit for thermostability for any structural class is not known, the potential thermostability of beetle luciferases was not known until demonstrations of the present invention.

*A priori* difficulties in achieving the goals of the present invention included:

1. The types of mutations which can be made by laboratory methods are limited.
  - i) By random point mutation (e.g. by error-prone PCR), more than one base change per codon is rare. Thus, most potential amino acid changes are rare.

ii) Other types of random genetic changes are difficult to achieve for areas greater than 100 bp (e.g., random gene deletions or insertions).

5           2. The number of possible luciferase mutants that can be screened is limited.

i) Based on sequence comparisons of natural luciferases, ignoring deletions and insertions, more than  $10^{189}$  functional enzyme sequences may be possible.

10           ii) If 100,000 clones could be screened per day, it would require more than  $10^{179}$  centuries to screen all possible mutants assuming same mutant was never screened twice (actual screening rate for the present invention was less than 5000 per day).

15           3. The probability of finding functional improvement requiring cooperative mutations is rare (the probability of finding a specific cooperative pair is 1 out of 108 clones).

Thus, even if the theoretical limits of thermostability were known, since only a very small number of the possible luciferase mutants can be screened, the *a priori* probability of finding such a thermostable enzyme was low.

20           However, the present invention now shows that it is possible and feasible to create novel beetle luciferases having high thermostability.

a) The approximately 250 mutants produced by methods of the present invention wherein the initial sequence was from *LucPpe2* and *LucPpe* demonstrate that it is possible and feasible for at least one member of this enzyme class to achieve high thermostability.

25           b) Any beetle luciferase should be improved by similar means since the luciferases belong to the same structural class.

30           i) Since all beetle luciferases belong to the same structural class, they also share in the same pool of potentially stabilizing mutations (this conclusion is supported by observation that a high percentage of the stabilizing mutations found in the clones of the present invention were

conversions to "consensus amino acids" in other beetle luciferases that is, amino acids that appear in the majority of beetle luciferase sequences (see FIG. 19).

- ii) Similar results were achieved using another beetle luciferase from the luminous beetle *Pyrophorus plagiophthalmus* (LucPplyG). The wild-type LucPplyG has 48% sequence identity to the wild type LucPpe2. Although the thermostability of the LucPplyG mutants were less than the LucPpe2 mutants described herein, this is because they were subjected to fewer cycles of directed evolution. Also, in some instances, mutants were selected with less emphasis placed on their relative thermostability. The most stable clone resulting from this evolution (Luc80-5E5) has a half-life of roughly 3.8 hours at 50°C.

To compensate for a statistical effect caused by the large number of deleterious random mutations expected relative to the beneficial mutations, methods were employed to maximize assay precision and to re-screen previously selected mutations in new permutations. Among the methods for maximizing assay precision were closely controlling culture conditions by using specialized media, reducing growth rates, controlling heat transfer, and analyzing parameters from mid-logarithmic phase growth of the culture, controlling mixing, heat transfers, and evaporation of samples in the robotic screening process; and normalizing data to spatially distributed control samples. New permutations of the selected mutations were created by a method of DNA shuffling using proof-reading polymerases.

The difficulty in predicting the outcome of the recursive process is exemplified by the variable success with the other characteristics of luciferase that were also selected for. Although the primary focus was on the enzyme thermostability, selection for mutants with brighter luminescence, more efficient substrate utilization, and an extended luminescence signal was also attempted. The definitions are given by equations herewith. The selection process was

determined by changes relative to the parent clones for each iteration of the recursive process. The amount of the change was whatever was observed during the screening process. The expression of luciferase in *E. coli* was relatively inefficient, for LucPpe2, compared to Luc+. Other luciferases varied (see Fig. 21).

To improve the overall efficiency of substrate utilization, reduction in the composite apparent utilization constant (i.e.,  $K_m$ -[ATP+luciferin]) for both luciferin and ATP was sought. Although there was an unexpected systematic change in each utilization constant, there was little overall change. Finally, the luminescence signal could only be moderately affected without substantially reducing enzyme efficiency. Thus, while the enzyme thermostability was greatly increased by methods of the present invention, other characteristics of the enzyme were much less affected.

FIGS. 48-53 present other results of the mutant luciferases. Compositions of the invention include luciferases having greater than the natural level of thermostability. Each mutant luciferase is novel, because its individual characteristics have not been reported. Specific luciferases are known by both their protein and gene sequences. Many other luciferases were isolated that have increased, high thermostability, but whose sequences are not known. These luciferases were identified during the directed evolution process, and were recognized as distinct by their enzymological characteristics.

A luciferase which is much more stable than any of the luciferase mutants previously described is designated as mutant Luc 90-1B5. New thermostable mutants were compared to this particularly stable luciferase. The mutant luciferases of the present invention display remarkable and heretofore unrealized thermostability at temperatures ranging from 22°C (room temperature) to at least as high as 65°C.

Other aspects of the invention include methods that incorporate the thermostable luciferases, specifically beetle luciferases having high thermostability.



### Production of Luciferases of the Present Invention

The method of making luciferases with increased thermostability is recursive mutagenesis followed by selection. Embodiments of the highly thermostable mutant luciferases of the invention were generated by a reiterative process of random point mutations beginning with a source nucleotide sequence, e.g. the cDNA *LucPpe2* [T249M] cDNA. Recombination mutagenesis is a part of the mutagenesis process, along with point mutagenesis. Both recombination mutagenesis and point mutagenesis are performed recursively. Because the mutation process causes recombination of individual mutants in a fashion similar to the recombination of genetic elements during sexual reproduction, the process is sometimes referred to as the sexual polymerase chain reaction (sPCR). See, for instance, Stemmer, U.S. Patent No. 5,605,793, issued February 25, 1997.

Taking the *LucPpe2* luciferase cDNA sequence as a starting point, the gene was mutated to yield mutant luciferases which are far more thermostable. A single point mutation to the *LucPpe2* sequence yielded the luciferase whose sequence is depicted as T249M. This mutant is approximately 5 times brighter *in vivo* than that of *LucPpe2*, it was utilized as a template for further mutation. It was also used as a baseline for measuring the thermostability of the other mutant luciferases described herein.

### Embodiments Of Sequences Of Luciferases Of The Present Invention

FIG. 45 shows the amino acid sequence of the *LucPpe2* luciferase. T249M. The sequence contains a single base pair mutation at position T249 to M (bold, underlined) which distinguishes it from the sequence reported by Leach *et al.*, (1997). This clone has a spectral maximum of 552 nm, which is yellow shifted from that of the Luc of Leach. This mutant was selected for use as an original template in some of the Examples because it is approximately 5 times brighter *in vivo*, than the form repeated by Leach *et al.* which allowed for more efficient screening by the assay. These sequences show changes from the starting sequence (T249-M) in bold face. Note that "x" in the sequence denotes an ambiguity in the sequence.

### Directed Evolution, A Recursive Process

Directed evolution is a recursive process of creating diversity through mutagenesis and screening for desired changes. For enzymological properties that result from the cumulative action of multiple amino acids, directed evolution provides a means to alter these properties. Each step of the process will typically produce small changes in enzyme function, but the cumulative effect of many rounds of this process can lead to substantial overall change.

The characteristic, "thermostability" is a candidate for directed evolution because it is determined by the combined action of many of the amino acids making up the enzyme structure. To increase the thermostability of luciferase, luminescence output and efficiency of substrate binding were also screened. This was to ensure that changes in thermostability did not also produce undesirable changes in other important enzymological properties.

Because the frequency of deleterious mutations is much greater than useful mutations, it is likely that undesirable clones are selected in each screen within the precision limits of the present invention. To compensate for this, the screening strategy incorporated multiple re-screens of the initially selected mutations. However, before re-screening, the selected mutations were "shuffled" to create a library of random intragenetic recombinations. This process allows beneficial mutations among different clones to be recombined together into fewer common coding sequences, and unlinks deleterious mutations to be segregated and omitted. Thus, although essentially the same set of selected mutations was screened again, they were screened under different permutations as a result of the recombination or shuffling.

Although results of each step of the evolutionary process were assayed by quantitative measurements, these measurements were mutually made in cell lysates rather than in purified enzymes. Furthermore, each step only measured changes in enzyme performance relative to the prior step, so global changes in enzyme function were difficult to judge. To evaluate the impact of directed evolution on enzyme function, clones from the beginning, middle and end of the

process (Table 2) were purified and analyzed. The clones selected for this analysis were Luc[T249M], 49-7C6, and 78-0B10. Another clone, 90-1B5, created by a subsequent strategy of oligonucleotide-directed mutagenesis and screening was also purified for analysis.

5           The effect of directed evolution on thermostability was dramatic. At high temperatures, where the parent clone was inactivated almost instantaneously, the mutant enzymes from the related clones showed stability over several hours (Table 1). Even at room temperature, these mutants are several fold more stable than the parent enzyme. Subsequent analysis of 90-1B5 showed this enzyme to be the  
10           most stable, having a half-life of 27 hours at 65°C when tested under the same buffer conditions. With some optimization of buffer conditions, this enzyme showed very little activity loss at 65°C over several hours (citrate buffer at pH 6.5; FIG. 1A). This luciferase was stable at room temperature over several weeks when incubated at pH 6.5 (FIG. 1B).

15           Kajiyama and Nakamo (1993) showed that firefly luciferase from *Luciola lateralis* was made more stable by the presence of a single amino acid substitution at position A217; to either I, L, or V. The substitution was from alanine. Substitution with leucine produced a luciferase that maintained 70% of its activity after incubation for 1 hour at 50°C. All of the enzymes of the present invention  
20           created through directed evolution, are much more stable than this *L. lateralis* mutant. The most stable clone, 90-1B5, maintains 75% activity after 120 hours (5 days) incubation under similar conditions (50°C, 25mol/L citrate pH 6.5, 150 mmol/L NaCl, 1mg/mL BSA, 0.1mmol/L EDTA, 5% glycerol). Interestingly, the Luc reported by Leach already contains isoleucine at the homologous position  
25           described for the *L. lateralis* mutant.

          Although thermostability was the characteristic of interest, clones were selected based on the other enzymological parameters in the screens. By selecting clones having greater luminescence expression, mutants were found that yielded greater luminescence intensity in colonies of *E. coli*. However, the process  
30           showed little ability to alter the kinetic profile of luminescence by the enzymes. This failure suggests that the ability to support steady-state luminescence is

integral to the catalytic mechanism, and is not readily influenced by a cumulative effect of many amino acids.

Substrate binding was screened by measuring an apparent composite  $k_m$  (see Example 2) for luciferin and ATP. Although the apparent composite  $K_m$  remained relatively constant, later analysis showed that the individual  $K_m$ 's systematically changed. The  $K_m$  for luciferin rose while the  $K_m$  for ATP declined (Table 2). The reason for this change is unknown, although it can be speculated that more efficient release of oxyluciferin or luciferin inhibitors could lead to more rapid enzyme turnover.

Each point mutation on its own increases (to a greater or lesser extent) the thermostability of the mutant enzyme beyond that of the wild-type luciferase. The cumulative effect of combining individual point mutations yields mutant luciferases whose thermostability is greatly increased from the wild-type, often on the order of a magnitude or more.

## EXAMPLES

The following examples illustrate the methods and compositions of the present invention and their embodiments.

### **EXAMPLE 1: Producing Thermostable Luciferases Of The Present Invention**

#### Mutagenesis Method:

An illustrative mutagenesis strategy is as follows:

From the "best" luciferase clone, that is a clone with improved thermostability and not appreciably diminished values for other parameters, random mutagenesis was performed by three variations of error-prone PCR. From each cycle of random mutagenesis, 18 of the best clones were selected. DNA was prepared from these clones yielding a total of 54 clones. These clones represent new genetic diversity.

These 54 clones were combined and recombination mutagenesis was performed. The 18 best clones from this population were selected.

These 18 clones were combined with the 18 clones of the previous population and recombination mutagenesis was performed. From this screening, a new luciferase population of 18 clones was selected representing 6 groups of functional properties.

In this screening the new mutations of the selected 54 clones, either in their original sequence configurations or in recombinants thereof, were screened a second time. Each mutation was analyzed on the average about 10 times. Of the 90 clones used in the recombination mutagenesis, it was likely that at least 10 were functionally equivalent to the best clone. Thus, the best clone or recombinants thereof should be screened at least 100 times. Since this was greater than the number of clones used in the recombination, there was significant likelihood of finding productive recombination of the best clone with other clones.

#### Robotic Processing Methods:

Heat transfers were controlled in the robot process by using thick aluminum at many positions where the 96-well plates were placed by the robotic arm. For example, all shelves in the incubators or refrigerator were constructed from ¼ inch aluminum. One position in particular, located at room temperature, was constructed from a block of aluminum of dimensions 4.5 x 7 x 6.5 inches. When any 96-well plate was moved from a high temperature (e.g., incubators) or low temperature (e.g., refrigerator) to a device at room temperature, it was first placed on the large aluminum block for temperature equilibration. By this means, the entire plate would rapidly reach the new temperature, thus minimizing unequal evaporation for the various wells in the plate due to temperature differences. Heat transfers in a stack of 96-well plates placed in an incubator (e.g., for overnight growth of *E. coli*) were controlled by placing 1 mm thick sheets of aluminum between the plates. This allowed for more efficient heat transfer from the edges of the stack to the center. Mixing in the robotic process was controlled by having the plate placed on a shaker for several second after each reagent addition.

Please refer to FIG. 14 for a schematic of the order in which the plates are analyzed (FIG. 15) and a robotic apparatus which can be programmed to perform the following functions:

5        Culture Dilution Method. A plate (with lid) containing cells is placed on a shaker and mixed for 3-5 minutes.

A plate (with lid) is gotten from a carousel and placed in the reagent dispenser. 180  $\mu$ l of media is added after removing the lid and placing on the locator near the pipetter. The plate is then placed in the pipetter.

10       The plate on the shaker is placed in the pipetter, and the lid removed and placed on the locator. Cells are transferred to the new plate using pipetting procedure (see "DILUTION OF CELLS INTO NEW CELL PLATE").

The lids are replaced onto both plates. The new plate is placed in the refrigerator and the old plate is returned to the carousel.

15       Luminescence Assay Method. A plate containing cells is retrieved from the carousel and placed on the shaker for 3-5 minutes to fully mix the cells. the cells tend to settle from solution upon standing.

To measure Optical Density (O.D.), the plate is moved from the shaker to the locator near the luminometer; the lid is removed and the plate placed into the  
20       luminometer. The O.D. is measured using a 620 nm filter.

When it is finished, the plate is then placed in the refrigerator for storage.

The above steps are completed for all plates before proceeding with subsequent processing.

To prepare a cell lysate, the plate of cells is first retrieved from the  
25       refrigerator and mixed on the shaker to resuspend the cells. A new plate from the carousel without a lid is placed in the reagent dispenser and 20  $\mu$ l of Buffer A is added to each well. This is placed in the pipetting station.

The plate of cells in the shaker is placed in the pipetting station. A daughter plate is prepared using pipetting procedure (see "PIPETTING CELLS  
30       INTO THE LYSIS PLATE") to prepare a daughter plate of cells.

After pipetting, the new daughter plate is placed on the shaker for mixing. The plate is returned to its original position in the carousel.

After mixing, the Lysate Plate is placed into the CO<sub>2</sub> freezer to freeze the samples. The plate is then moved to the thaw block to thaw for 10 minutes.

5           The plate is then moved to the reagent dispenser to add 175 µl of Buffer B, and then mixed on the shaker for about 15 minutes or more. The combination of the freeze/thaw and Buffer B will cause the cells to lyse.

10           A new plate with a lid from the carousel is used to prepare the dilution plate from which all assays will be derived. The plate is placed in the reagent dispenser and the lid removed to the locator near the pipetter. 285 µl of Buffer C is added to each well with the reagent dispenser, then the plate is placed in the pipetting station.

15           The Lysate Plate in the shaker is moved to the pipetting station and pipetting procedure (see "DILUTION FROM LYSIS PLATE TO INCUBATION PLATE") is used. After pipetting, the new daughter plate is placed on the shaker for mixing. The Lysate Plate is discarded.

20           Two white assay plates are obtained from the plate feeder and placed in the pipetter. The incubation plate from the shaker is placed in the pipetter, and the lid removed and placed on the nearby locator. Two daughter plates are made using the pipetting procedure (see CREATE PAIR OF DAUGHTER PLATES FROM INCUBATION PLATE"). Afterwards, the lid is replaced on the parent plate, and the plate is placed in a high temperature incubator. [ranging from 31° to about 65° depending on the clone.]

25           One daughter plate is placed in the luminometer and the 1x ASSAY METHOD is used. After the assay, the plate is placed in the ambient incubator, and the second daughter plate is placed in the luminometer. For the second plate, the 0.02x ASSAY METHOD is used. This plate is discarded, and the first plate is returned from the incubator to the luminometer. The REPEAT ASSAY method is used (i.e., no reagent is injected). Afterwards, the plate is again returned to the  
30           ambient incubator.

The above steps are completed for all plates before proceeding with processing.

To begin the second set of measurements, the plate from the high temperature incubator is placed in the shaker to mix.

5           The plate in the ambient incubator is returned to the luminometer and the REPEAT ASSAY method is again used. The plate is returned afterwards to the ambient incubator.

10           Two white assay plates again are obtained from the plate feeder and placed in the pipetter. The plate on the shaker is placed in the pipetter, and the lid removed and placed on the nearby locator. Two daughter plates are again made using the pipetting procedure (see "CREATE PAIR OF DAUGHTER PLATES FROM INCUBATION PLATE"). Afterwards, the lid is replaced on the parent plate, and the plate is returned to the high temperature incubator.

15           One daughter plate is placed in the luminometer and the 1x ASSAY METHOD is again used. The plate is discarded after the assay. The second daughter plate is then placed in the luminometer and the 0.06x ASSAY METHOD is used. This plate is also discarded.

The above steps are completed for all plates before proceeding with processing.

20           In the final set of measurements, the plate from the high temperature incubator is again placed in the shaker to mix.

The plate in the ambient incubator is returned to the luminometer and the REPEAT ASSAY method is again used. The plate is discarded afterwards.

25           One white assay plate is gotten from the plate feeder and placed in the pipetter. The plate from the shaker is placed in the pipetter, and the lid removed and placed on the nearby locator. One daughter plate is made using the pipetting procedure (see "CREATE SINGLE DAUGHTER PLATE FROM INCUBATION PLATE"). The lid is replaced on the parent plate and the plate is discarded.

30           The daughter plate is placed in the luminometer and the 1x ASSAY METHOD is used. The plate is discarded after the assay.



Buffers:Buffer A:

5           25mM K<sub>2</sub>HPO<sub>4</sub>  
            .5mM CDTA  
            .1% Triton X-100

Buffer B:

10          X CCLR (Promega e153a)  
            1.25mg/ml lysozyme  
            0.04% gelatin

Buffer C:

15          10mM HEPES  
            150mM NaCl  
            1mg/ml BSA  
            5% glycerol  
            0.1mM EDTA  
  
20          1X Assay reagent:  
            5uM Luciferin  
            175uM ATP  
            20mM Tricine , pH 8.0  
            0.1mM EDTA  
  
25          0.02X Assay reagent:  
            1:50 dilution of 1X Assay reagent  
            0.06X Assay reagent:  
            1:150 dilution of 1X Assay reagent

Pipetting Procedures  
Pipetting Cells Into the Lysis Plate

Non-aseptic procedure using fixed tips

On the pipetter deck:

- 5                                -place a plate containing approximately 200  $\mu$ l cells without lid  
                                     -Lysate Plate containing 20  $\mu$ l of Buffer A

Procedure:

1. Move the tips to the washing station and wash with 1 ml.
2. Move to the cell plate and withdraw 60  $\mu$ l.
- 10                                3. Move to the Lysate Plate and dispense 45  $\mu$ l.
4. Repeat steps 1-3 for all 96 samples.
5. At the conclusion of the procedure, step 1 is repeated to clean the tips.

Post-procedure:

- Place Lysate Plate onto the shaker.
- 15                                - Place lid on plate with cells and place on carousel.
- Place Lysate Plate into the CO<sub>2</sub> freezer.

DILUTION FROM LYSIS PLATE TO INCUBATION PLATE

20                                On the pipetter deck:

- Lysate Plate containing 240  $\mu$ l of lysate  
                                     - Incubation Plate without lid containing 285  $\mu$ l of Buffer C

Procedure:

1. Move the tips to the washing station and wash with 0.5 ml.
- 25                                2. Move to the Lysate Plate and withdraw 30  $\mu$ l.
3. Move to the Incubation Plate and dispense 15  $\mu$ l by direct contact with the buffer solution.
4. Repeat steps 1-3 for all 96 samples.
5. At the conclusion of the procedure, step 1 is repeated to clean the tips.

## Post-procedure:

- Place Incubation Plate on shaker.
- Discard Lysate Plate.

## 5 CREATE PAIR OF DAUGHTER PLATES FROM INCUBATION PLATE

This procedure is done twice

On the pipetter deck:

- Incubation Plate containing 100-300 µl of solution without lid
- Two empty Assay Plates (white)

10

Procedure:

1. Move the tips to the washing station and wash with 0.5 ml.
2. Move to the Incubation Plate and withdraw 50 µl.
3. Move to the first Assay Plate and dispense 20 µl.
4. Move to the second Assay Plate and dispense 20 µl.
- 15 5. Repeat steps 1-4 for all 96 samples.
6. At the conclusion of the procedure, step 1 is repeated to clean the tips.

## Post-procedure:

1. Replace lid on Incubation Plate.
2. Place Incubation Plate in incubator.
- 20 3. Place first Assay Plate in luminometer.
4. Place second Assay Plate on carousel.

## CREATE SINGLE DAUGHTER PLATE FROM INCUBATION PLATE

25 On the pipetter deck:

Place incubation Plate containing 100-300 µl of solution without lid  
and  
Empty Assay Plate (white)

Procedure:

1. Move the tips to the washing station and wash with 0.5 ml.
2. Move to the Incubation Plate and withdraw 40  $\mu$ l.
3. Move to the Assay Plate and dispense 20  $\mu$ l.
- 5 4. Repeat steps 1-3 for all 96 samples.
5. At the conclusion of the procedure, step 1 is repeated to clean the tips.

Post-procedure:

- Discard Incubation Plate and lid on Incubation Plate.
- Place Assay Plate in luminometer.

10

**DILUTION OF CELLS INTO NEW CELL PLATE**

Aseptic procedure using fixed tips

On the pipetter deck:

- plate containing approximately 200  $\mu$ l of cells without lid
- 15 - new cell plate containing 180  $\mu$ l of Growth Medium without lid

Procedure:

1. Move to the cell plate and withdraw 45  $\mu$ l.
2. Move to the Cell Plate and dispense 20  $\mu$ l volume by direct liquid-to-liquid transfer.
- 20 3. Move to waste reservoir and expel excess cells.
4. Move to isopropanol wash station aspirate isopropanol to sterilize tips.
5. Move to wash station, expel isopropanol and wash tips.
6. Repeat steps 1-4 for all 96 samples.

Post-procedure:

- 25 1. Replace lid on original plate of cells and place onto carousel.
2. Replace lid on new cell plate and place into refrigerator.

Notes:

This procedure is used to prepare the cell plates used in the main analysis procedure.

180 µl of Growth Medium is added by the reagent dispenser to each of the new cell plates just prior to initiating the pipetting procedure.

The dispenser is flushed with 75% isopropanol before priming with medium.

5           The medium also contains selective antibiotics to reduce potential contamination.

#### Luminometer Procedures

#### 1x ASSAY METHOD

10

- place plate into luminometer

1. Inject 100 ul of 1X Assay reagent
2. Measure luminescence for 1 to 3 seconds
- 15 3. Repeat for next well

- continue until all wells are measured

20

#### 0.02x ASSAY METHOD

- place plate into luminometer

1. Inject 100 ul of 0.02X Assay reagent
- 25 2. Measure luminescence for 1 to 3 seconds
3. Repeat for next well

- continue until all wells are measured

30

#### 0.06x ASSAY METHOD

- place plate into luminometer

- 35 1. Inject 100 ul of 0.06X Assay reagent
2. Measure luminescence for 1 to 3 seconds
3. Repeat for next well

- continue until all wells are measured

40

REPEAT ASSAY

- place plate into luminometer

- 5      1. Measure luminescence for 1 to 3 seconds  
2. Repeat for next well

- continue until all wells are measured

#### *IN VIVO* SELECTION METHOD

10            5-7 nitrocellulose disks, 200-500 colonies per disk (1000-3500 colonies total), are screened per 2 microplates (176 clones). The clones are screened at high temperatures using standard screening conditions.

15            8 positions in each microplate are reserved from a reference clone using the "best" luciferase (the parent clone for random mutagenesis and codon mutagenesis). The positions of the reserved wells is shown as "X" below.

	XooooooooX
	oooooooooooo
20	oooXooooXooo
	oooooooooooo
	oooooooooooo
	oooXooooXooo
	oooooooooooo
25	XooooooooX

30            The reference clones are made by placing colonies from DNA transformed from the parent clone into the reference wells. (To identify these wells prior to inoculation of the microplate, the wells are marked with a black marking pen on the bottom of each well).

Wood and

## SCREENING SELECTION CRITERIA

The following were used to screen. Criteria 1 is achieved manually; data for criteria 2-6 is generated by robotic analysis. For all criteria, the maximum value as described are selected.

- 5                   1. ***In vivo* screen.** The brightest clones are selected at an elevated temperature.
2. **Expression/specific activity.** The value of normalized luminescence are calculated as the ratio of luminescence to optical density. The values are reported as the ratio with the reference value.
- 10               3. **Enzyme stability.** Measurements of normalized luminescence of the incubated samples (3 taken over about 15 hours) are fitted to  $\ln(L)=\ln(L_0)-(t/\tau)$ , where L is normalized luminescence and t is time.  $\tau$  is a measure of the enzyme stability. The value is reported as the ratio with the reference value, and the correlation coefficients are
- 15               calculated.
4. **Substrate binding.** Measurements of normalized luminescence with 1x and 0.02x are taken at the initial reading set, and 1x and 0.06x are taken at the 5 hour set. The ratio of the 0.02x:1x and 0.06x:1x gives the relative luminescence at 0.02x and 0.06x concentrations. These
- 20               values, along with the relative luminescence at 1x (i.e., 1), are fitted to a Lineweaver-Burk plot to yield the  $K_m:app, total$  for the substrates ATP, luciferin, and CoA. The value are reported as the inverse ratio with the reference value, and the correlation coefficients are
- calculated.
- 25               5. **Signal stability.** The luminescence of the initial 1x luminescent reaction are re-measured 3 additional times over about 15 hours. These values are fitted to  $\ln(L)=\ln(L_0)-(t/\tau)$  and the integral over t (15 hours) are calculated. Signal stability is then calculated as  $S=(1-\int(L)/L_0t)^2$ . The value are reported as the inverse ratio with the
- 30               reference value, and the correlation coefficient are calculated.

6. **Composite fitness.** The values of criteria 2 through 5 are combined into a single composite value of fitness (or commercial utility). This value is based on a judgment of the relative importance of the other criteria. This judgment is given below:

5

<u>Criteria</u>	<u>Relative Value</u>
Stability	5
Signal Stability	2
Substrate Binding	2
Expression/Activity	1

10

The composite,  $C = \text{Sum}(\text{criteria 2-5 weighted by relative value, e.g., more weight is on stability because that was a major goal})$ .

#### EXAMPLE 2: Software

**Procedure: Organize data into SQL database.** Each file created by a luminometer (96 well) (Anthos, Austria) represents the data from one microplate. These files are stored in the computer controlling the luminometer, and connected to the database computer by a network link. From each microplate of samples, nine microplates are read by the luminometer (the original microplate for optical density and eight daughter microplates for luminescence).

Ninety files are created in total; each containing data sets for 96 samples. Each data set contains the sample number, time of each measurement relative to the first measurement of the plate, luminometer reading, and background corrected luminometer reading. Other file header information is also given. The time that each microplate is read is also needed for analysis. This can be obtained from the robot log or the file creation time. A naming convention for the files are used by the robot during file creation that can be recognized by SQL (e.g. YYMMDDPR.DAT where YY is the year, MM is the month, DD is the day, P is the initial plate [0-9], and R is the reading [0-8]).



**Procedure: Data Reduction And Organization.**

- 5      - **Normalize luminescence data:** For each measurement of luminescence in the eight daughter plates, the normalized luminescence is calculated by dividing by the optical density of the original plate. If any value of normalized luminescence is less than zero, assign the value of  $0.1sL$  where  $sL$  is the standard deviation for measurements of normalized luminescence.
- 10     - **Calculate relative measurement time:** For each normalized luminescence measurement, the time of the measurement is calculated relative to the first measurement of the sample. For example, the time of all luminescence measurements of sample B6 in plate 7 (i.e., 7:B06) are calculated relative to the first reading of 7:B06. This time calculation will involve both the time when the plate is read and the relative time of when the sample is read in the plate.
- 15     - **Calculate enzyme stability ( $\tau$ ):** For each sample, use linear regression to fit  $\ln(L_{1x}) = \ln(L_0) - (t/\tau)$  using the three luminescence measurements with  $1x$  substrate concentrations (Plates 1, 5, 8). Also calculate the regression coefficient.
- 20     - **Calculate substrate binding ( $K_{m:app,total}$ ):** Using microplates from the first set of readings (Plates 1 and 2), calculate the  $L_{0.02x,rel}$  by dividing measurements made with substrate concentrations of  $0.02x$  by those of  $1x$ . Similarly, calculate the  $L_{0.06x,rel}$  using microplates of the second set of readings (Plates 5 and 6), by dividing measurements made with substrate concentrations of  $0.06x$  by those of  $1x$ .

For each sample, use linear regression to fit  $1/L = (K_{m:app,total}/L_{max:app}) (1/[S]) + (1/L_{max:app})$  using

	L	[S]
25	$L_{0.02x,rel}$	0.02
	$L_{0.06x,rel}$	0.06
	1 ( $L_{1x,rel}$ )	1

$K_{m:app,total}$  is calculated as the slope/intercept. Also calculate the regression coefficient.

30

- **Calculate signal stability (S):** For each sample, use linear regression to fit  $\ln(L)=\ln(L_0)-(t/\tau)$  using the four luminescence measurements of the initial microplate with 1x substrate concentrations (Plates 1, 3, 4, and 7). Also calculate the regression coefficient. From the calculated values of  $\tau$  and  $L_0$ , calculate the integral of luminescence by  $\text{int}(L)=\tau L_0 (1-\exp(-t_f/\tau))$ , where  $t_f$  is the average time of the last measurement (e.g., 15 hours). The signal stability is calculated as  $S=(1-\text{int}(L)/L_i t_f)^2$ , where  $L_i$  is the initial measurement of normalized luminescence with 1x substrate concentration (Plate 1)

[Note: To correct for evaporation, an equation  $S=(1+K-\text{int}(L)/L_i t_f)^2$ , may be used where  $1/K=2(\text{relative change of liquid volume at } t_f)$ .]

- **Calculate the reference value surfaces:** A three dimensional coordinate system can be defined by the using the grid positions of the samples within a microplate as the horizontal coordinates, and the calculated values for the samples ( $L_i$ ,  $K_{m:\text{app},\text{total}}$ ,  $\tau$ , or S) as the vertical coordinate. This three dimensional system is referred to as a "plate map". A smooth surface in the plate maps representing a reference level can be determined by least squares fit of the values determined for the 8 reference clones in each microplate. For each of the 10 initial microplates of samples, respective reference surfaces are determined for the criteria parameters  $L_i$ ,  $\tau$ ,  $K_{m:\text{app},\text{total}}$ , and S (40 surfaces total).

In the least squares fit, the vertical coordinate (i.e., the criteria parameters) are the dependent variable, the horizontal coordinates are the independent variables. A first order surface (i.e.,  $z=ax+by+c$ ) are fitted to the values of the reference clones. After the surface is calculated, the residuals to each reference clone are calculated. If any of these residuals is outside of a given cutoff range, the reference surface are recalculated with omission of the aberrant reference clone.

If a first order surface does not sufficiently represent the values of the reference clones, a restricted second order surface are used (i.e.,  $z=a(x^2+ky^2)+bx+cy+d$ , where k is a constant).

- 5       - **Calculate the reference-normalized values:** For the criteria parameter of each sample, a reference-normalized values is determined by calculating the ratio or inverse ratio with the respective reference value. The reference-normalized values are  $L_i/L_{ir}$ ,  $\tau/\tau_r$ ,  $K_{mr}/K_{m:app,total}$ , and  $S_r/S$ , where reference values are calculated from the equations of the appropriate reference surface.
- **Calculate the composite scores:** For each sample, calculate  $C=5(\tau/\tau_r)+2(S_r/S)+2(K_{mr}/K_{m:app,total})+(L_i/L_{ir})$ .
- 10       - **Determine subgroupings:** For the criteria parameters  $L_i$ ,  $\tau$ ,  $K_{m:app,total}$ ,  $S$ , and  $C$ , delimiting values (i.e., bin sizes) for subgroupings are defined as  $gL$ ,  $g\tau$ ,  $gK_m$ ,  $gS$ , and  $gC$ . Starting with the highest values for  $L_i$ ,  $\tau$ , or  $C$ , or the lowest values of  $K_{m:app,total}$  or  $S$ , the samples are assigned to bins for each criteria parameter (the first bin being #1, and so on).
- **Display sorted table of reference-normalized values:** Present a table of data for each sample showing in each row the following data:
- 15               - sample identification number (e.g., 7:B06)
- composite score ( $C$ )
- reference-normalized enzyme stability ( $\tau/\tau_r$ )
- correlation coefficient for enzyme stability
- bin number for enzyme stability
- 20               - reference-normalized signal stability ( $S_r/S$ )
- correlation coefficient for signal stability
- bin number for signal stability
- reference-normalized substrate binding ( $K_{mr}/K_{m:app,total}$ )
- correlation coefficient for substrate binding
- 25               - bin number for substrate binding
- reference-normalized expression/specific activity ( $L_i/L_{ir}$ )
- bin number for expression/specific activity

The table is sorted by the composite score ( $C$ ).

**Procedure: Present sorted table of criteria parameters.**

Present a table of data for each sample showing in each row the following data:

- sample identification number
- composite score (C)
- 5      - enzyme stability ( $\tau$ )
- correlation coefficient for enzyme stability
- bin number for enzyme stability
- signal stability (S)
- correlation coefficient for signal stability
- 10     - bin number for signal stability
- substrate binding ( $K_{m:app,total}$ )
- correlation coefficient for substrate binding
- bin number for substrate binding
- expression/specific activity ( $L_i$ )
- 15     - bin number for expression/specific activity

The table is sorted by the composite score (C); the reference clones are excluded from the table. Same entry coding by standard deviation as described above.

**Procedure: Present sorted table of reference-normalized values.**

- 20      This is the same procedure as the final step of the data reduction procedure. The table will show:

- sample identification number
- composite score (C)
- reference-normalized enzyme stability ( $\tau/\tau_r$ )
- 25     - correlation coefficient for enzyme stability
- bin number for enzyme stability
- reference-normalized signal stability ( $S_r/S$ )
- correlation coefficient for signal stability
- bin number for signal stability

- reference-normalized substrate binding ( $K_{m,r}/K_{m:app,total}$ )
- correlation coefficient for substrate binding
- bin number for substrate binding
- reference-normalized expression/specific activity ( $L_i/L_{ir}$ )
- 5      - bin number for expression/specific activity

The table is sorted by the composite score (C); the reference clones are excluded from the table. Same entry coding by standard deviation as described above.

**Procedure: Present sorted table of criteria parameters for reference clones.**

10      This is the same procedure as described above for criteria parameters, except for only the reference clones. The table will show:

- sample identification number
- composite score (C)
- enzyme stability ( $\tau$ )
- 15      - correlation coefficient for enzyme stability
- bin number for enzyme stability
- signal stability (S)
- correlation coefficient for signal stability
- bin number for signal stability
- 20      - substrate binding ( $K_{m:app,total}$ )
- correlation coefficient for substrate binding
- bin number for substrate binding
- expression/specific activity ( $L_i$ )
- bin number for expression/specific activity

25      The table is sorted by the composite score (C). Same entry coding by standard deviation as described above.

**Procedure: Present sorted table of reference-normalized values.**

This is the same procedure as described above for reference-normalized values, except for only the reference clones. The table will show:

- sample identification number
- composite score (C)
- reference-normalized enzyme stability ( $\tau / \tau_r$ )
- correlation coefficient for enzyme stability
- 5      - bin number for enzyme stability
- reference-normalized signal stability ( $S_r/S$ )
- correlation coefficient for signal stability
- bin number for signal stability
- reference-normalized substrate binding ( $K_{mr}/K_{m:app,total}$ )
- 10      - correlation coefficient for substrate binding
- bin number for substrate binding
- reference-normalized expression/specific activity ( $L_i/L_{ir}$ )
- bin number for expression/specific activity

15      The table is sorted by the composite score (C). Same entry coding by standard deviation as described above.

**Procedure: Sort table.**

Any table may be sorted by any entries as primary and secondary key.

**Procedure: Display histogram of table.**

20      For any table, a histogram of criteria parameter vs. bin number may be displayed for any criteria parameter.

**Procedure: Display plate map.**

For any plate, a plate map may be displayed showing a choice of:

- any luminescence or optical density measurement
- $L_i$
- 25      -  $L_i$  reference surface
- $L_i/L_{ir}$
- $\tau$
- $\tau$  reference surface

- $\tau / \tau_r$
- correlation coefficient of  $\tau$
- S
- S reference surface
- 5 -  $S_r/S$
- correlation coefficient of S
- $K_{m:app,total}$
- $K_m$  reference surface
- $K_{mr}/K_{m:app,total}$
- 10 - correlation coefficient for  $K_{m:app,total}$
- composite score (C)

The plate maps are displayed as a three dimensional bar chart. Preferably, the bars representing the reference clones are indicated by color or some other means.

**Procedure: Display drill-down summary of each entry.**

- 15 For  $L_i$ ,  $\tau$ ,  $K_{m:app,total}$ , and S, any entry value in a table may be selected to display the luminescence and optical density reading underlying the value calculation, and a graphical representation of the curve fit where appropriate. Preferably the equations involved and the final result and correlation coefficient will also be display.
- 20 -  $L_i$  or  $L_i/L_r$ . Display the optical density and luminescence value from the chosen sample in Plate 0 and Plate 1.
- $\tau$  or  $\tau / \tau_r$ . Display the optical density and luminescence value from the chosen sample in Plate 0, Plate 1, Plate 5, and Plate 8. Display graph of  $\ln(L/x)$  vs. t, showing data points and best line.
- 25 - S or  $S_r/S$ . Display the optical density and luminescence value from the chosen sample in Plate 0, Plate 1, Plate 3, Plate 4, and Plate 7. Display graph of  $\ln(L)$  vs. t, showing data points and best line.
- $K_{m:app,total}$  or  $K_{mr}/K_{m:app,total}$ . Display the optical density and luminescence value from the chosen sample in Plate 0, Plate 1, Plate 2, Plate 5, and Plate 6.
- 30 Display graph of  $1/L$  vs.  $1/[S]$ , showing data points and best line.

**EXAMPLE 3: Preparation Of Novel Luciferases**

The gene with FIG. 1 contains a single base pair mutation at position 249, T to M. This clone has a spectral maximum of 552 nm which is yellow shifted from the sequence of Luc. This mutant was selected as an original template because it is about 5 time brighter *in vivo* which allowed for more efficient screening.

C-terminus mutagenesis

To eliminate the peroxisome targeting signal (-SKL) the L was mutated to a STOP and the 3 codons immediately upstream were randomized according to the oligonucleotide mutagenesis procedure described herein. The mutagenic oligonucleotide designed to accomplish this also introduces a unique SpeI site to allow mutant identification without sequencing. The mutants were screened *in vivo* and 13 colonies picked, 12 of which contained the SpeI site.

N-terminus mutagenesis

To test if expression could be improved, the 3 codons immediately downstream from the initiation Met were randomized as described herein. The mutagenic oligo designed to accomplish this also introduces a unique ApaI site to allow mutant identification without sequencing. Seven clones were selected, and six of the isolated plasmids were confirmed to be mutants.

Shuffling of C- and N-terminus mutants

The C- and N-terminus mutagenesis was performed side-by-side. To combine the N and C-terminus mutations, selected clones from each mutagenesis experiment were combined with the use of recombination mutagenesis according to the recombination mutagenesis protocol described herein. The shuffled mutants were subcloned into amp<sup>r</sup> pRAM backbone and screened in DH5 F'IQ. [BRL, Hanahan, 1985) A total of 24 clones were picked, only 4 contained both the N- and C- terminus mutations. These 4 clones were used as templates for randomization of the cysteine positions in the gene.



**Mutagenesis to randomize cysteine positions/Random mutagenesis and recombination mutagenesis in the Luc gene**

There are 7 cysteine positions in the Ppe-2 gene. It is known that these positions are susceptible to oxidation which could cause destabilization of the protein. Seven oligonucleotides were ordered to randomize the cysteine positions.

The oligonucleotides were organized into two groups based upon the conservation of cysteine in other luciferase genes from different families. Group 1 randomizes the conserved cysteine positions C-60, C-80, and C-162. Group 2 randomizes cysteines that are not strictly conserved at positions C-38, C-127, C-221, and C-257.

The four selected templates from the N and C terminus mutagenesis were sub-cloned into an ampicillin-sensitive backbone and single-stranded DNA was prepared for each of the templates. These templates were combined in equal amounts and oligonucleotide mutagenesis was completed as described herein. It was determined by plating an aliquot of the mutS transformation prior to overnight incubation that each of the 2 groups contained  $2 \times 10^4$  independent transformants. MutS-DNA was prepared for the 2 groups and was then transformed into JM109 cells for screening. Mutants from group 1 were screened *in vivo* and picks were made for a full robotic run. Five clones were selected that had improved characteristics. Mutants from group 2 were screened *in vivo* and picks were made for a full robotic run. The temperature incubator on the robot was set at 33°C for this set of experiments. Ten clones were selected that had improved characteristics.

The fifteen best picks from both groups of the cysteine mutagenesis experiments were shuffled together as described herein and 18 of the best clones were selected after robotic processing.

The "best" clone from the above experiment (31-1G8) was selected as a template for subsequent rounds of mutagenesis. (The high temperature robot incubator temperature was set to 42°C) Another complete round of mutagenesis was completed.

The 18 best clones from the above mutagenesis were picked and clone (39-5B10) was selected as the best clone and was used as a template for another round of mutagenesis. (The high temperature robot incubator temperature was set at 49°C).

5           After this cycle, 6 of the best clones were selected for sequencing. Based upon the sequence data, nine positions were selected for randomization and seven oligos were designed to cover these positions. Based upon data generated from the robot, it was determined that the best clone from the group of six clones that were sequenced was clone (49-7C6). The luciferase gene from this clone was sub-  
10           cloned into an ampicillin-sensitive pRAM backbone and single stranded DNA was prepared. The randomization of the selected positions was completed according to the oligonucleotide mutagenesis procedure listed above.

          The randomization oligos were divided into 4 groups, and transformants from these experiments were picked and two robotic runs were completed. Ten  
15           clones were selected from the two experiments. (The high temperature robot incubator temperature on robot was set at 56°C).

          The best 10 picks from the above two experiments, and the best 18 picks from the previous population of clones were shuffled together (recombination mutagenesis protocol).

20           The 18 best clones were selected and clone 58-0A5 was determined to be the best clone. This clone was then used as a template for another round of mutagenesis. The high temperature robot incubator temperature was set at 56°C. Clone 71-504 was selected as a new lead clone and another round of mutagenesis was completed. Incubator set at 60°C.

25           The best 18 picks were selected and the best clone from this group was determined to be clone 78-0B10. The temperature stability of clones at various temperatures is presented in the FIGS.

#### **EXAMPLE 4: Mutagenesis Strategy From Clone 78-0B10 to 90-1B5**

1. 23 oligos (oligonucleotides) were ordered to change 28 positions to consensus. All of the oligos were tested individually using oligo directed mutagenesis with single stranded DNA from clone *luc78-0B10* as a template to determine which oligos gave an improvement in stability. Below is a table which lists the mutagenic oligos.

5

**Description****OLIGO SYNTHESIS  
NUMBER**

A17 to T	6215
M25 to L	6216
S36 to P; remove <i>Nsi</i> I site	6217
A101 to V, S105 to N	6218
I125 to V	6219
K139 to Q	6220
V145 to I	6221
V194 to I	6222
V203 to L, S204 to P	6231
A216 to V	6232
A229 to Q	6233
<b>M249 to T (reversion)</b>	<b>6234</b>
T266 to R, K270 to E	6235
E301 to D	6236
N333 to P, F334 to G	6237
R356 to K	6238
I363 to V	6246
A393 to P	6247
R417 to H	6248
G482 to V	6249
N492 to T	6250
F499 to Y, S501 to A	6251
L517 to V	6252
F537 to L	6253

\*Note that oligo #6234 does not change a consensus position. This oligo causes a reversion of position 249 to the wild-type PPE-2 codon. Although reversion of this position was shown to increase thermostability, reversion of this position decreased light output.

10

2. Oligonucleotide-directed mutagenesis with clone *luc78-0B10* as a template:

Based on the results of individually testing the mutagenic oligonucleotides listed above, three experiments were completed and oligos for these experiments were divided in the following manner:

a. 6215,6234,6236,6248 (found to give increased stability)

b.

215,6217,6218,6219,6220,6221,6222,6231,6233,6234,6236,6238,6247,6248,6249,6251,6253.

(found to be neutral or have increased stability.)

c. All 23 oligos.

3. Selections from the three experiments listed above were screened with the robotic screening procedure (Experiment 84). (*luc78-0B10* used as a control).

4. Selections from experiment 84 were recombined using the recombination mutagenesis procedure and then screened with the robotic screening procedure (Experiment 85).

5. Single stranded DNA was prepared from three (3) clones, *luc85-3E12*, *luc85-4F12*, *luc85-5A4*. These clones were used as templates for oligonucleotide-directed mutagenesis to improve codon usage. Positions were selected based upon a codon usage table published in Nucleic Acids Research vol. 18 (supplement) 1990. page. 2402. The table below lists oligos that were used to improve codon usage in *E. coli*.

Description	Oligo synthesis #
L7-(tta-ctg), remove <i>Apa</i> I site	6258
L29-(tta-ctg)	6259
T42-(aca-acc)	6260
L51,L56-(tta-ctg),L58-(ttg-ctg)	6261
L71-(tta-ctg)	6262
L85-(ttg-ctg)	6263
L95-(ttg-ctg),L97(ctt-ctg)	6273

L113,L117-(tta-ctg)	6274
L151,L153-(tta-ctg)	6275
L163-(ctc-ctg)	6276
R187-(cga-cgt)	6277
L237-(tta-ctg)	6279
R260-(cga-cgc)	6280
L285,L290-(tta-ctg),L286-(ctt-ctg)	6281
L308-(tta-ctg)	6282
L318-(tta-ctg)	6283
L341-(tta-ctg),T342-(aca-acc)	6284
L380-(ttg-ctg)	6285
L439-(tta-ctg)	6286
L456-(ctc-ctg),L457-(tta-ctg)	6293
T506-(aca-acc),L510-(cta-ctg)	6305
R530-(aga-cgt)	6306

6. In the first experiment, the three templates listed above from Experiment 85 were combined and used as a templates for oligonucleotide-directed mutagenesis. All of the oligos were combined in one experiment and clones resulting from oligonucleotide-directed mutagenesis were screened using the robotic screening procedure as Experiment 88. There were a low percentage of luminescent colonies that resulted from this experiment, so another oligonucleotide-directed mutagenesis experiment was completed in which the oligonucleotides were combined in the following groups:

- a. 6258,6273,6280,6286
- b. 6259,6274,6281,6293
- c. 6260,6275,6282,6294
- d. 6261,6276,6283,6305
- e. 6262,6277,6284,9306
- f. 6263,6279,6285

7. It was discovered that samples from group b had a low amount of luminescent colonies, and it was hypothesized that one of the oligos in group b was causing problems. Selections were made from all of the experiments with the exception of experiment b. Samples were then run through the robotic screening procedure (Experiment 89).
8. Selections from Experiments 88 and 89 were shuffled together with the recombination mutagenesis protocol and were then screened with the robotic screening procedure (Experiment 90).

## MATERIALS AND METHODS

### A. Mutagenesis Protocol

The mutant luciferases disclosed herein were produced via random mutagenesis with subsequent *in vivo* screening of the mutated genes for a plurality of characteristics including light output and thermostability of the encoded luciferase gene product. The mutagenesis was achieved by generally following a three-step method:

1. **Creating genetic diversity through random mutagenesis.** Here, error-prone PCR of a starting sequence such as that of Luc was used to create point mutations in the nucleotide sequence. Because error-prone PCR yields almost exclusively single point mutations in a DNA sequence, a theoretical maximum of 7 amino acid changes are possible per nucleotide mutation. In practice, however, approximately 6.1 amino acid changes per nucleotide is achievable. For the 550 amino acids in luciferase, approximately 3300 mutants are possible through point mutagenesis.
2. **Consolidating single point mutations through recombination mutagenesis.** The genetic diversity created by the initial mutagenesis is recombined into a smaller number of clones by sPCR. This process not only reduces the number of mutant clones, but because the rate of mutagenesis is high, the probability of linkage to negative mutations is significant. Recombination mutagenesis unlinks positive mutations from

negative mutations. The mutations are "re-linked" into new genes by recombination mutagenesis to yield the new permutations. Then, after re-screening the recombination mutants, the genetic permutations that have the "negative mutations" are eliminated by not being selected.

5 Recombination mutagenesis also serves as a secondary screen of the initial mutants prepared by error-prone PCR.

3. **Broadening genetic diversity through random mutagenesis of selected**

**codons.** Because random point mutagenesis can only achieve a limited number of amino acid substitutions, complete randomization of selected  
10 codons is achieved by oligonucleotides mutagenesis. The codons to be mutated are selected from the results of the preceding mutagenesis processes on the assumption that for any given beneficial substitution, other alternative amino acid substitutions at the same positions may produce even greater benefits. The positions to be mutated are identified by DNA  
15 sequencing of selected clones.

**B. Initial mutagenesis experiments**

Both the N-terminus and the C-terminus of the starting sequence were modified by oligonucleotide-directed mutagenesis to optimize expression and remove the peroxisomal targeting sequence. At the N-terminus, nine bases  
20 downstream of the initiation **CODON** were randomized at the C-terminus, nine bases upstream of the termination **CODON** were randomized. Mutants were analyzed using an *in vivo* screen, resulting in no significant change in expression.

Six clones from this screen were pooled, and used to mutate the codons for seven cysteines. These codons were randomized using oligonucleotide-directed  
25 mutagenesis, and the mutants were screened using the robotic screening procedure. From this screen, fifteen clones were selected for directed evolution.

**C. Generating and Testing Clones**

Several very powerful and widely known protocols are used to generate and test the clones of the present invention. Unless noted otherwise, these  
30 laboratory procedures are well known to one of skill in the art. Particularly noted

as being well known to the skilled practitioner is the polymerase chain reaction (PCR) devised by Mullis and various modifications to the standard PCR protocol (error-prone PCR, sPCR, and the like), DNA sequencing by any method (Sanger's or Maxam & Gilbert's methodology), amino acid sequencing by any method  
5 (e.g., the Edman degradation), and electrophoretic separation of polynucleotides and polypeptides/proteins.

#### **D. Vector Design**

A preferred vector (pRAM) used for the mutagenesis procedure contains several unique features that allow for the mutagenesis strategy to work efficiently:  
10 The pRAM vector contains a filamentous phage origin, *f1*, which is necessary for the production of single-stranded DNA.

Two *Sfi*I sites flank the gene. These sites were designed by so that the gene to be subcloned can only be inserted in the proper orientation.

The vector contains a *tac* promoter.  
15 Templates to be used for oligonucleotide mutagenesis contain a 4 base-pair deletion in the *bla* gene which makes the vector ampicillin-sensitive. The oligonucleotide mutagenesis procedure uses a mutant oligonucleotide as well as an ampicillin repair oligonucleotide that restores function to the *bla* gene. This allows for the selection of a high percentage of mutants. (If selection is not used,  
20 it is difficult to obtain a high percentage of mutants.)

#### **E. Uses of Luciferases**

The mutant luciferases of the present invention are suitable for use in any application for which previously known luciferases were used, including the  
25 following:

ATP Assays. The greater enzyme stability means that reagents designed for detection of ATP have a greater shelf-life and operational-life at higher temperatures (e.g., room temperature). Therefore, a method of detecting ATP using luciferases with increased thermostability, is novel and useful.



Luminescent labels for nucleic acids, proteins, or other molecules.

Analogous to advantages of the luciferases of the present invention for ATP assays, their greater shelf-life and operational-life is a benefit to the reliability and reproducibility of luminescent labels. This is particularly advantageous for labeling nucleic acids in hybridization procedures where hybridization temperatures can be relatively high (e.g. greater than 40°C. Therefore, a method of labeling nucleic acids, proteins, or other molecules using luciferases of the present invention is novel and useful.

Genetic reporter. In the widespread application of luciferase as a genetic reporter, where detection of the reporter is used to infer the presence of another gene or process of interest, the increased thermal stability of the luciferases provides less temperature dependence of its expression in living cells and in cell-free translations and transcription/translation systems. Therefore a method using the luciferases of the present invention, as genetic reporters is novel and useful.

Enzyme immobilization. Enzymes in close proximity to physical surfaces can be denatured by their interaction with that surface. The high density immobilization of luciferases onto a surface to provide strong localized luminescence is improved by using high stability luciferases. Therefore, a method of immobilizing luciferases onto a solid surface using luciferases of the present invention, is novel and useful.

Hybrid proteins. Hybrid proteins made by genetic fusion genes encoding luciferases and of other genes, or through a chemical coupling process, benefit by having a greater shelf-life and operational-life. Therefore, a method of producing hybrid proteins through genetic means or chemical coupling using the luciferases of the present invention, is novel and useful.

High temperature reactions. The light intensity of a luciferase reaction increases with temperature until the luciferase begins to denature. Because the use of thermostable luciferases allows for use at greater reaction temperatures, the luciferases of the present invention are novel and useful for performing high temperature reactions.

Luminescent solutions. Luminescence has many general uses, including educational, demonstrational, and entertainment purposes. These applications benefit from having enzymes with greater shelf-life and operational-life. Therefore, a method of making luminescent solutions using the luciferases of the present invention, is novel and useful.

#### **F. Firefly luciferase**

The firefly luciferase gene chosen for directed evolution was Luc isolated from *Photuris pennsylvanica*. The luciferase was cloned from fireflies collected in Maryland by Wood *et al.* and later was independently cloned by Dr. Leach using fireflies collected in Oklahoma (Ye *et al.*) (1977). A mutant of this luciferase (T249M) was made by Wood *et al.* and used in the present invention because it produced approximately 5-fold more light when expressed in colonies of *E. coli*.

Overview of Evolution Process: Directed evolution was achieved through a recursive process, each step consisting of multiple cycles of 1) creating mutational libraries of firefly luciferase followed by 2) screening the libraries to identify new mutant clones having a plurality of desired enzymological characteristics.

To begin the process, three mutational libraries were created using error-prone PCR (Fromant *et al.*, 1995). Each library was screened first by visual evaluation of luminescence in colonies of *E. coli* (Wood and De Luca, 1987), and then by quantitative measurements of enzymological properties in *E. coli* cell lysates. Approximately 10,000 colonies were examined in the visual screen, from which 704 were selected for quantitative analysis. From each quantitative screen 18 clones were selected.

The three sets of 18 clones each were pooled together, and a new mutational library was created using DNA shuffling to generate intragenetic recombinations (sPCR; Stemmer, 1994). The results were screened to yield another set of 18 clones. The entire process was completed by combining this set of 18 clones with 18 clones from the previous round of evolution, creating another mutational library by DNA shuffling, and screening as before.

Screening method: In the qualitative visual screen, colonies were selected only for their ability to sustain relatively bright luminescence. The thermal stability of the luciferase within the colonies of *E. coli* was progressively challenged in successive rounds of evolution by increasing the temperature of the screen. The selected colonies were inoculated into wells of 96-well plates each containing 200µl of growth medium.

In the quantitative screens, lysates of the *E. coli* cultures were measured for 1) luminescence activity, 2) enzyme stability, 3) sustained enzymatic turnover, and 4) substrate binding.

“Luminescence activity” was measured as the ratio of luminescence intensity to the optical density of the cell culture.

“Enzyme stability” was determined by the rate of activity loss from cell lysates over 10 hours. In successive rounds of evolution the incubation temperature of the lysates was increased.

“Sustained enzymatic turnover” was determined by the rate of luminescence loss of a signal enzymatic reaction over 10 hours at room temperature. “Substrate binding” was determined by the relative activity of the lysate when assayed with diluted substrate mixtures. Of these four parameters, the highest priority for selection was placed on thermostability.

Robotic Automation. Robotic automation was used in the quantitative screens to accurately perform the large number of required quantitative assays on the cultured cells. Overnight cultures were first diluted into fresh medium and grown for 3 hours to produce cultures in mid-log phase growth. The optical densities of each cultures was then measured, and aliquots of the cultures were lysed by freeze/thaw and lysozyme. The resulting lysates were further diluted before analysis and incubated at elevated temperatures. Luminescence was measured from aliquots of the diluted lysates, taken at various times, and measured under various conditions as prescribed by the analytical method (see Example 2). Computer analysis of this data yielded the quantitative selection criteria described above.

Summary of evolutionary progression: After mutagenesis of the N- and C-termini, and randomization of the cysteine codons, a pool of 15 clones was subjected to two rounds of directed evolution as described herein. Five of the 18 clones resulting from this process were sequenced to identify mutations. One of these clones designated, 49-7C6, was chosen for more detailed analysis and further mutagenesis. This clone contained 10 new amino acid substitutions compared to the luciferase Luc[T249M].

To assess the potential for other amino acid replacements at the sites of these substitutions, oligonucleotide-directed mutagenesis was used to randomize these codons. The resulting clones were screened as described herein, and 18 selected clones were used to initiate two new rounds of directed evolution. Of the 18 clones resulting from this second set of rounds, the clone designated 78-0B10 was chosen for additional study and mutagenesis. This clone encoded a luciferase that contained 16 new amino acid substitutions compared to Luc[T249M].

Using oligonucleotide directed mutagenesis with 78-0B10 as the template, codons were selected for substitution to consensus amino acids previously known among beetle luciferases. Selections from this mutagenesis experiment were shuffled together and three clones, determined to be the most stable were then used as templates for oligonucleotide mutagenesis to improve codon usage in *E. coli*. A clone designated 90-1B5 selected from this experiment, contained 28 amino acid substitutions relative to Luc[T249M]. Out of 25 codons selected for change to consensus amino acids, 11 were replaced in the clone designated 90-1B5. Only five out of the 30 positions that were selected for improved codon usage were substituted and had little effect on enzyme expression.

**Protein purification** The four mutants that are described herein (Luc[T249M], 49-7C6, 78-0B10, and 90-1B5) were purified using a previously published procedure (Hastings *et al.*, 1996).

**Enzymological characterization** Purified proteins were diluted in 25mmol/L HEPES pH 7.8, 150mmol/L NaCl, 0.1mmol/L EDTA, 1mg/mL BSA. Enzyme stability was determined from diluted proteins incubated at different temperatures, and aliquots were removed at different time points. A linear

regression of the natural log of the luminescence and time was calculated.  
Half-life was calculated as the  $\ln(0.5)/\text{slope}$  of the regression.

#### E. PCR Mutagenesis Protocol (Random Mutagenesis):

##### *PCR mutagenesis reactions*

1. Prepare plasmid DNA from a vector containing the gene of interest, estimate DNA concentration from a gel.
2. Set up two 50  $\mu\text{l}$  reaction reactions per group:

There are three groups of mutagenic conditions using different skewed nucleotide concentrations.

The conditions listed herein yield in the range of from 8-10% wild-type Luc colonies after subcloning phenotypic for each generated parent clone. The rate of mutagenesis is estimated by the number of luminescent colonies that are present after mutagenesis. Based upon results of clones mutated in the range of 8-10%, it was determined that this level of mutagenesis produces on average approximately 2-3 amino acid changes per gene. If the mutagenesis rate is selected so that on average there is one amino acid change per gene, then on average 50% of the clones will have no mutations. (Bowie, *et al.*, 1990).

For the master mix: add all components except polymerase, vortex, spin briefly, add polymerase, and mix gently.

Component	AtoT/TtoA	AtoC/TtoG	Gtoa/CtoT
Datp	0.3mM	0.1mM	0.25mM
Dctp	2.75mM	4mM	1mM
DGTP	0.06mM	0.02mM	0.05mM
DTTP	0.625mM	0.3mM	0.6mM
*pRAMtailUP	0.4 pmol/ul	0.4 pmol/ul	0.4 pmol/ul
*pRAMtailDN	0.4 pmol/ul	0.4 pmol/ul	0.4 pmol/ul
*Taq. Polymerase	1U/ul	1U/ul	1U/ul
$^{\circ}\text{MgCl}_2$	6.77mM	5.12mM	2.7mM
$^{\circ}\text{MnCl}_2$	0.5mM	0.5mM	0.3mM
DNA	50ng total	50ng total	50ng total

Component	AtoT/TtoA	AtoC/TtoG	GtoA/CtoT
10x PCR buffer	1X	1X	1X
Autoclaved nanopure	To 50 ul	To 50 ul	To 50 ul

\* Taq. Polymerase is purchased from Perkin Elmer (N808-0101).

10x Taq polymerase buffer (aliquot the Taq into 1.5 ml tubes and store at -70°C):

- 100mM Tris-HCl pH8.4 from 1M stock

5 - 500mM KCL

Primers are diluted from a 1 nmol/ $\mu$ l stock to a 20 pmol/ $\mu$ l working stock.

pRAMtailup: 5'-gtactgagacgacgccagcccaagcttaggcctgagtg-3'

pRAMtaildn: 5'-ggcatgagcgtgaactgactgaactagcgccgccgag-3'

10 ° MnCl<sub>2</sub> and MgCl<sub>2</sub> are made fresh from 1M stocks. The stocks are filter sterilized and mixed with sterile water to make the 10mM and 25mM stocks which are then stored in Polystyrene Nalgene containers at 4°C.

Cycle in thermal cycler: 94°C for 1min (94°C-1min, 72°C-10min) 10x.

3. Purify reaction products with Wizard PCR purification kit (Promega Corporation, Madison, Wisconsin, part#A718c):

15 - transfer PCR reaction into a new tube containing Promega 100  $\mu$ l

Direct Purification buffer (Part#A724a)

- add 1 ml of Wizard PCR Purification Resin (part#A718c) Promega and incubate at room temperature for 1min

- pull resin through Wizard minicolumn

20 - wash with 80% Ethanol

spin in microcentrifuge to remove excess Ethanol

- elute into 50  $\mu$ l sterile nanopure water (allow water to remain on column for at least 1 min)

*Amplification<sup>1</sup> Of Mutagenesis Reaction*

1. Set up five 50 ml reactions per group:
  - To master mix: add all components except polymerase, vortex, spin briefly, add polymerase, mix gently.
- 5            ° 10x reaction buffer for Native PFU contains 20mM MgCl<sub>2</sub>, so no additional MgCl<sub>2</sub> needs to be added
  - + primers:
    - pRAM18up -5'gtactgagacgacgccag-3'
    - pRAM19dn -5'ggcatgagcgtgaactgac-3'
- 10           Cycling conditions: 94-30 sec (94-20 sec, 65-1 min, 72-3 min) 25x (Perkin-Elmer Gene Amp® PCR System 2400)
  2. Load 1 µl on a gel to check amplification products
  3. Purify amplification reaction products with Wizard PCR purification kit (Promega Corporation, part#A718c):
    - 15            - transfer PCR reaction into a new tube containing 100 µl Direct Purification buffer (Promega, Part#A724a)
    - add 1 ml of Wizard PCR Purification Resin (Promega Part#A718c) and incubate at room temperature for 1 min
    - pull resin through Wizard minicolumn
    - 20            - wash with 80% Ethanol
    - spin in microcentrifuge to remove excess Ethanol
    - elute with 88 µl sterile nanopure water (allow water to remain on column for at least 1 min)

---

<sup>1</sup> This amplification step with PFU Polymerase was incorporated for 2 reasons:  
 (a) To increase DNA yields for the production of large numbers of transformants.  
 (b) To reduce the amount of template DNA that is carried over from the mutagenic PCR reaction: (Primers for the second amplification reaction are nested within the mutagenic primers. The mutagenic primers were designed with non-specific tails of 11 and 12 bases respectively for the upstream and downstream primers. The nested primers will amplify DNA that was previously amplified with the mutagenic primers, but cannot amplify pRAM template DNA.)

***Subcloning of amplified PCR mutagenesis products***

1. Digest the DNA with *Sfi*I as follows:
  - 2  $\mu$ l *Sfi*I (Promega Part #R639a)
  - 10  $\mu$ l 10X buffer B (Promega Part #R002a)
  - 5       - 88  $\mu$ l of DNA from Wizard PCR prep (see step 3 [in amplification])
  - mix components and overlay with 2 drops of mineral oil; incubate at 50°C for 1 hour
2. Remove salts and *Sfi*I ends with Wizard PCR purification as described  
10       herein, and  
      elute into 50  $\mu$ l sterile nanopure water
3. Ligation into pRAM (+/r) backbone (set up 4 ligations per group):
  - 0.025 pmol pRAM backbone
  - 0.05 pmol insert (usually in the range of 6 to 12  $\mu$ l of insert)
  - 15       - 1  $\mu$ l of T4 DNA Ligase (M180a)
  - 2  $\mu$ l of 10x ligase buffer (C126b, divide into 25  $\mu$ l aliquots, do not freeze/thaw more than twice)
  - water to 20  $\mu$ l
  - ligate for 2 hours at room temperature
  - 20       - heat reactions for 15 min at 70 C to inactivate ligase

***Transformation and plating***

1. Butanol precipitate samples to remove excess salts (n-Butanol from Sigma, St. Louis, Missouri, part #BT-105):
  - (if Ethanol precipitation is used instead of butanol a wash with 70%  
25       ethanol as needed) (excess salt will cause arcing during the electroporation which causes the reaction to fail)
  - add water to 50  $\mu$ l
  - add 500  $\mu$ l of n-butanol
  - mix until butanol/ligation mix is clear and then spin for 20 min at  
30       room temperature



- drain butanol into waste container in fume hood
- resuspend in 12  $\mu$ l water, spin 30 sec at full speed
- 2. Preparation of cell/DNA mix (set up 4 transformations plus one with reference clone DNA):
  - 5                   - while DNA is precipitating, place electroporation cuvettes on ice
  - fill 15 ml Falcon snap-cap tubes with 3 ml S.O.C. medium and place on ice
  - thaw JM109 electrocompetent cells on ice (50  $\mu$ l per ligation reaction)
  - 10               - pipette 10  $\mu$ l of the bottom layer from step 1 (or 0.5  $\mu$ l ref. clone DNA) into competent cells
  - (small amounts of butanol carry-over do not adversely effect the transformation efficiency)
  - place cell/DNA mix on ice
  - 15               3. Electroporation:
    - carry tubes, cuvettes, and cell/DNA mix on ice to electroporation device
    - pipette cell-DNA mix into a cuvette and zap. Instrument settings:
      - Cuvette gap: 0.2 cm
      - 20           Voltage:               2.5 kV
      - Capacitance: 25  $\mu$ F
      - Resistance: 200 Ohms
      - Time constant:       4.5 msec
    - pipette 1 ml SOC (contains KCL; media prep #KCLM) into
    - 25           cuvette, quickly pour into recovery tube (transformation efficiency is reduced if cells are allowed to sit in cuvette)
    - place the recovery tube on ice until all samples are processed
    - allow the cells to recover at 37°C for 30-60 min
    - plate on LB+amp plates with nitrocellulose filters
    - 30           (# of colonies is ~20% higher if cells recover 60 min, possibly due to cell replication. See 101305 p.65)

(Best colony density for screening is 500 per plate. For the current batch of cells plate ~500 to 750µl)

**F. Recombination Mutagenesis Protocol or DNA shuffling:**

***DNase I digestion of plasmid DNA***

- 5           1. Prepare 2% low melting point gel
  - use 0.8g agarose in 40 ml (NuSieve #50082)
  - use large prep comb
  - make sure it is solidified prior to digesting
2. Prepare 4 µg of pooled plasmid DNA for digest
- 10       3. Prepare 1 U/µl DNase dilution on ice according to the table below:

Dnase I <sup>+</sup>	0.74 µl
10x DnaseI buffer	10 µl
1% gelatin*	10 µl
Water to 100 µl	

\* DNase I from Sigma (D5791)

\* Gelatin was added to keep the DNase I from sticking to the walls of the tubes.

15           This dilution can be kept on ice for at least 30 min without loss in activity.

4. Digest (set up at room temperature):
  - prepare two digests with 1.0U and 1.5U DNaseI per 100 µl reaction:
  - 10 µl of 10x DNase I buffer (500mM Tris, 10mM MgCl<sub>2</sub> pH 7.8)
  - x µl DNA ( 2µg of pooled plasmid DNA from step 2)
  - 20       - 1 or 1.5 µl of the 1U/µl enzyme dilution
  - sterile nanopure water to 100 µl
  - incubate at room temperature for 10 minutes
  - stop reaction with 1µl of 100mM CDTA

***Purification from agarose gel***

- 25       1. Run DNase digested fragments on gel

- add 10 µl of 10x blue juice to each DNase I digest
- load all on a 2% Low melting point agarose gel
- run about 30 min at 120-150V
- load pGEM DNA marker in middle lane

5

## 2. Isolate fragments

- cut out agarose slice containing fragments in the size range of 600-1000bp using a razor blade

- cut into pieces that weigh ~0.3g

- melt the gel slices at 70°C

10

- add 300 µl of Phenol (NaCl/Tris equilibrated) to the melted agarose, vortex for ~1 min at max speed

- spin for 10 min at 4°C (the interface is less likely to move around if it is done at 4°C)

15

- remove the top layer into a tube containing an equal volume of Phenol/Chloroform/Isoamyl (saturated with 300mM NaCl /100mM Tris pH 8.0), vortex and spin for 5 min at RT

- remove the top layer into a tube containing chloroform and vortex and spin.

20

- remove the top layer into a tube with 2 vol. of 95% cold Ethanol; place in -70°C freezer for 10 min (no additional salts are needed because of the High Salt Phenol)

- spin at 4°C for 15 minutes.

- wash with 70% Ethanol, drain and air dry for ~10 min

- resuspend in 25 to 50 µl of sterile nanopure water

25

- store at -70°C until ready for use

### *Assembly reaction*

Set up 4 reactions and pool when completed

Component	Concentration	Amount in $\mu$ l	Final concentration
dATP	10 mM	1	200 $\mu$ M
dCTP	10 mM	1	200 $\mu$ M
dGTP	10 mM	1	200 $\mu$ M
dTTP	10 mM	1	200 $\mu$ M
DNA*		5	
Tli	3U/ $\mu$ l	0.4	0.24 U/ $\mu$ l
10X Thermo buffer	10X	5	1X
MgCl <sub>2</sub>	25mM	4	2mM
gelatin	1%	5	0.1%
water		To 50 $\mu$ l	

\* Because the DNA used for this reaction has been fragmented, it is difficult to estimate a concentration. The easiest way is to load 5  $\mu$ l of the DNaseI digested DNA to an agarose gel and run the gel until the dye enters the wells (1-2 min). Fragments from a typical 2 $\mu$ g DNA digest which were resuspended in 100  $\mu$ l of water will give a DNA concentration of ~1 to 10 ng/ $\mu$ l. See 101284 p.30 for a photo of this type of gel.

Cycling conditions: 94-30sec [94-20sec, 65-1min, 72-2min] 25x (Program "assembly-65", runs ~2.5 h)

#### *Amplification of assembly*

Usually 5 amplification reactions will produce enough DNA for a full 8 plate robotic run

Component	Concentration	Amount in $\mu$ l	Final concentration
Datp	10 mM	1	200 $\mu$ M
dCTP	10 mM	1	200 $\mu$ M
dGTP	10 mM	1	200 $\mu$ M
dTTP	10 mM	1	200 $\mu$ M
pRAMtailup*	20 pmol/ $\mu$ l	2	0.8 pmol/ $\mu$ l
pRAMtaildn*	20 pmol/ $\mu$ l	2	0.8 pmol/ $\mu$ l
PFU native polymerase <sup>+</sup>	2 U/ $\mu$ l	1	0.04 U/ $\mu$ l
10x native PFU buffer <sup>o</sup>	1x	5	1x

DNA		5	
water		water to 50 $\mu$ l	

\* Note that the concentration of primers is twice as high as in a typical amplification reaction.

° The PFU 10X buffer contains 20mM MgCl<sub>2</sub>, so it is not necessary to add MgCl<sub>2</sub>.

+ PFU is ordered from Stratagene part #600135.

Cycling conditions: 94-30sec [94-20sec, 65-1min, 72-3min] 25x

### *Subcloning of assembly amplification*

#### 1. Purify amplification products with Wizard PCR purification:

- pool 5 amplification reactions
- transfer into a new tube that contains 100  $\mu$ l of Direct Purification buffer
- add 1 ml of Wizard PCR Purification Resin, incubate at RT for 1 min
- pull Resin through Wizard minicolumn
- wash with 80% ethanol and spin in microcentrifuge to remove excess ethanol
- elute with 88  $\mu$ l of sterile nanopure water (allow water to remain on column for at least 1 min)

#### 2. Digest with SfiI:

- 2  $\mu$ l SfiI
- 10  $\mu$ l 10x buffer B
- 88  $\mu$ l of DNA from Wizard PCR prep
- mix components and overlay with 2 drops of mineral oil; incubate at 50°C for 1 hour

#### 3. Band isolation:

Sometimes after amplification of the assembly reaction a band that is smaller than the gene-sized fragment is produced. This small fragment has been

shown to subclone about 10-fold more frequently than the gene sized fragment if the sample is not band isolated. When this contaminating band is present, it is necessary to band isolate after Sfi I digestion.

- load the DNA to a 0.7% agarose gel
- 5        - band isolate and purify with the Gene Clean kit from Bio 101
- elute DNA with 50 µl sterile nanopure water, check concentration on gel (This type of purification with standard agarose produced the highest number of transformants after subcloning. Other methods tried: Low melt with Phenol chloroform, Gene clean with low melt, Wizard PCR resin with standard
- 10       agarose, Pierce Xtreme spin column with Low melt (did not work with standard agarose)).
- 4. Ligate into pRAM [+/-] backbone: (See ligation and transformation protocol above)

***Large scale preparation of pRAM backbone***

- 15       1. Streak an LB amp plate with pRAMMCS [+/-] (This vector contains a synthetic insert with a SacII site in place of a gene. It can be found in -70°C in box listed pRAM glycerol stocks position b2. This vector contains the new ribosome binding site, but it will be cut out when the vector is digested with SfiI.
- 20       2. Prepare a 10 ml overnight culture in LB supplemented with amp.
- 3. The next day inoculate 1L of LB supplemented with amp and grow for 16-20 hours.
- 4. Purify the DNA with the Wizard Maxi Prep kit. (use 4 preps for 1L of cells)
- 25       5. Digest the Plasmid with SfiI. (Use 5U per microgram) Overlay with mineral oil and digest for at least two hours.
- 6. Ethanol Precipitate to remove salts. Resuspend in water.
- 7. Digest with SacII for 2 hours. (keep digest volume to 2 ml or less).
- 30       It is possible that part of the plasmid could be partially digested. If the vector is cut with an enzyme that is internal to the two SfiI sites, it will

keep the partially digested fragments from joining in a ligation reaction.

8. Load entire digest onto a column (see 9). The volume of the sample load should not be more than 2 ml. If it is it will be necessary to ethanol precipitate.
9. The column contains Sephacryl s-1000 and is stored with 20% ethanol to prevent bacterial contamination. Prior to loading the sample the column must be equilibrated with cold running buffer for at least 24 hours. If the column has been sitting more than a couple of months it may be necessary to empty the column, equilibrate the resin 3-4 washes in cold running buffer, and then re-pour the column. After the column is poured it should be equilibrated overnight so that the resin is completely packed.
10. Collect fractions of ~0.5ml. Typically the DNA comes off between fractions 25 and 50. Load a five  $\mu$ l aliquot from a range of fractions to determine which fractions contain the backbone fragment. The small insert fragment will start to come off the column before all of the backbone is eluted, so it will be necessary to be conservative when fractions are pooled. For this reason typically 40-60% of the DNA is lost at this step.
11. Pool the fractions that contain the backbone.
12. Ethanol precipitate the samples. Resuspend in a volume that produces ~10-50 ng/  $\mu$ l.
13. Store at -70°C.

Column running buffer: (store at 4°C)

5 mM EDTA

100 mM NaCl

50 mM Tris-HCL pH 8.0

10  $\mu$ g/ml tRNA (R-8759)

**H. Oligonucleotide Mutagenesis:**

Prepare Ampicillin-sensitive Single stranded DNA of the template to be mutated.  
Design a mutagenic primer that will randomly generate all possible amino acid  
codons.

*Mutagenesis reaction:*

Component	Final concentration
Single Stranded Template	0.05pmol
Mutagenic Oligo	1.25pmol
Ampicillin Repair Oligo (Promega q631a)	0.25pmol
10X annealing buffer	1X
Water to 20 ul	

\*Annealing buffer:

-200mM Tris-HCl, pH 7.5

-100mM MgCl<sub>2</sub>

-500mM NaCl

Heat reaction at 60°C for 15 minutes and then immediately place on ice.

Synthesis reaction:

Component	Amount
Water	5 ul
10X synthesis buffer	3 ul
T4 DNA Polymerase (Promega m421a)	1 ul (10 Units)
T4 DNA Ligase (Promega 180a)	1 ul (3 Units)

\*Synthesis buffer

100mM Tris-HCl, pH 7.5

5mM dNTPs

10mM ATP

20mM DTT



Incubate at 37C for 90 minutes.

Transform into Mut-S strain BMH 71-18 (Promega strain Q6321)

-Place Synthesis reaction in a 17X100mm tube.

5       -Add BMH 71-18 competent cells that have been thawed on ice to  
      synthesis reaction.

      -Incubate on ice for 30 min

      -Heat Shock cells at 42°C for 90 seconds.

      -Add 4 ml of LB medium and grow cells at 37C for 1 hour. Add  
Ampicillin to a final concentration of 1.25ug/ml and then grow overnight at 37°C.

10       Isolate DNA with Wizard Plus Purification system (Promega a7100)

      Transform isolated DNA into JM109 electro-competent cells and transform  
      onto LB Ampicillin plates.

#### I.     Screening procedure:

      JM109 clones (from a transformation reaction) are plated onto  
15   nitrocellulose filters placed on LB amp plates at a screening density of ~500  
      colonies per plate.

      As listed in the Random Mutagenesis procedure, approximately 10% of the  
      clones to be selected will have to be as stable as the same sequenced or better than  
      source. Or stated another way, ~50 colonies per plate will be suitable for  
20   selection. There are 704 wells available for a full eight plate robotic run, so at  
      least 15 LB amp plates will be needed for a full robotic run.

      After overnight growth at 37°C the plates contains the transformants are  
      removed from the incubator and placed at room temperature.

      The nitrocellulose filter is lifted on one side and 500 µl of 10mM IPTG is  
25   added to each of the plates. The filter is then placed back onto the plate to allow  
      diffusion of the IPTG into the colonies containing the different mutant luciferase  
      genes. The plates are then incubated for about 4 hours at room temperature.

      One (1) ml of a solution contains 1mM Luciferin and 100mM Sodium  
      Citrate is pipetted onto a slide warmer that is set at 50°C. A nitrocellulose filter  
30   that contains mutant luciferase colonies and has been treated with IPTG is then

placed on top of the luciferin solution. After several minutes, the brightest colonies are picked with tooth picks which are used to inoculate wells in a microtiter plate that contain M9- minimal media with 1% gelatin.

5 After enough colonies are picked to 8 microtiter plates, the plates are placed in an incubator at 350rpm at 30°C incubation and are grown overnight.

In the morning the overnight plates are loaded onto the robot and the cell dilution procedure is run. (This procedure dilutes the cultures 1:10 into induction medium). The new plates are grown for 3 hours at 350rpm at 30°C.

10 After growth, the plates are loaded to the robot for the main assay procedure.

**Minimal Media:**

6g/Liter Na<sub>2</sub>HPO<sub>4</sub>  
 3g/Liter KH<sub>2</sub>PO<sub>4</sub>  
 0.5g/Liter NaCl  
 15 1g/Liter NH<sub>4</sub>Cl  
 2mM MgSO<sub>4</sub>  
 0.1mM  
 1mM Thiamine-HCl  
 0.2% glucose  
 20 12ug/ml Tetracycline  
 100ug/ml ampicillin

\*Overnight media contains 1% gelatin

\*Induction media contains 1mM IPTG and no gelatin.

25 S.O.C. Media  
*-10mM NaCl*  
*-2.5mM KCl*  
*-20mM MgCl*  
*-20mM glucose*

*-2% bactotryptone*

*-0.5% yeast extract*

**TABLE 1: Parameters Characterizing Luciferases of Clones Derived for Various Experiments**

Control is  
PPE-2 39-  
5B10 at 51C.

Experiment	Clone ID	Li	tau	Km	S
40	0a7	1.04	4.5	0.78	1
40	5h4	1.29	1.61	1.16	0.953
40	0c2	1.13	1.54	0.91	0.998
40	5g4	1	1.4	0.85	1
40	6d3	1.02	1.37	0.79	1
40	1g4	1.06	1.28	0.77	0.985
40	1d4	1.69	1.23	0.73	1
40	0h9	1.26	1.21	0.63	0.998
40	2f6	3	1.07	0.49	0.981
40	7d6	3.09	1.058	1.09	1.013
40	5a7	4.3	1.025	0.93	1.008
40	4c8	1	1	0.33	1.004

Experiment	Clone ID	Li	tau	Km	S
41	7h7	0.73	2.4	2.1	0.995
41	5a5	0.77	1.93	2.7	1.002
41	2c12	1.06	1.7	0.91	1.003
41	6e5-	1.16	1.62	1.53	0.997
41	4e5-	1.08	1.37	1.4	1.004
41	6g7	1.3	1.27	1.39	0.999
41	1h4	1.36	1.24	0.56	0.994
41	0c11	4.1	1.23	1.24	0.996
41	2h9	5.3	1.01	0.83	0.986
42	6b10	0.97	3.6	0.97	0.997
42	1c3	0.91	2.1	0.6	0.998
42	7h9	0.8	1.8	0.8	0.982
42	6b2	0.77	1.72	0.8	0.978
42	6d6	0.83	1.7	0.733	0.975
42	4e10-	0.77	1.63	1.8	0.954
42	1b5	0.83	1.41	1.05	0.955
42	6e6-	0.71	1.16	0.89	0.955
42	3a9	0.85	1.3	0.86	0.997
42	6b6	2.7	1.3	0.91	1.02

42	6e9-	1.5	1.27	0.98	1.01
42	3h11	1.73	1.21	0.63	0.985
42	1a2	1.11	1.17	0.77	1.005
42	3f7	0.49	1.16	1.13	0.944
42	1a4	2	1.01	0.76	0.996

Control is  
PPE-2 40-  
0A7 at 54C

Experiment	Clone ID	Li	tau	Km	S
46	2h3	0.86	6.4	0.37	0.96
46	4a9	0.67	5.7	0.66	0.997
46	2g4	0.65	5.3	0.78	0.96
46	5d12	0.94	4.9	0.94	1.002
46	1h11	1.02	4.8	0.84	0.998
46	5a10	1.23	4.4	0.81	0.9842
46	0a8	1.35	4.3	0.89	1
46	4d3	0.51	3.6	0.65	0.975
46	2a3	1.17	2.9	0.57	0.988
46	3b11	1.39	2.5	0.63	1.02
46	7g12	1.49	2.5	0.91	1.02
46	0g9	1.86	2.25	0.5	0.998
46	7h8	1.07	1.36	0.52	0.99
46	1g8	0.3	1.31	0.72	0.92
46	1d3	1.74	1.13	1.02	1.001
46	0c3	1.68	1.01	0.74	1.01
46	5c11	0.82	1.01	0.6	0.95

Control is  
PPE-2 46-  
2h3 at 54.

Experiment	Clone ID	Li	tau	Km	S
49	6c10	0.57	2.2	0.98	1
49	7c6	1.12	1.9	0.93	1.01
49	0g12	1	1.58	0.69	1.08
49	7a5	1.08	1.44	1.1	0.99
49	1f6	0.66	1.13	1.04	1.006
49	0b5	0.76	1.07	1.03	0.98
49	4a3	0.94	1.06	0.77	1

Control is  
PPE-2 49-  
7C6 at 56C  
Experiment

Experiment	Clone ID	Li	tau	Km	S
56	2d12	0.97	2.9	0.29	1.006
56	5g10	1.01	2.77	0.64	1.007
56	3d5	1.32	2.25	1.85	1.03

Experiment	Clone ID	Li	tau	Km	S
57	3d1	1.06	2.9	1.05	1.02
57	6g12	1	2.7	0.87	1.004
57	4c1	0.79	2.6	0.93	1.014
57	5f10	0.72	1.9	0.64	1.03
57	1e6-	0.84	1.49	0.984	0.9871
57	1h2	0.94	1.43	0.68	0.991
57	2a6	1.08	1.08	0.89	0.9976

Experiment	Clone ID	Li	tau	Km	S
58	1g6	1.57	8.9	1.78	1.02
58	0a5	1.53	8.5	1.56	1.05
58	1b1	0.84	8.5	0.6	1.04
58	3g1	1	7.34	0.62	1.006
58	0f3	1.31	6.9	0.57	0.98
58	3e12-	1.06	6.3	0.47	0.996
58	0c7	1.9	4	0.64	1.06
58	0d1	1.03	3.76	0.49	1.03
58	3c7	1.49	3.4	0.55	1.04
58	2a2	1.4	2.2	0.5	1.05
58	2a8	3.2	2	0.81	1.05
58	0f2	2.2	1.92	0.45	1.04
58	1b4	5.1	1.87	1.08	1.09
58	2b3	2.7	1.55	0.57	1.04
58	4g1	4.9	1.2	0.72	1.06

Control is  
PPE-2 58-  
0A5 at 58C

Experiment	Clone ID	Li	tau	Km	S
61	4e9-	1.03	1.84	0.76	1.01
61	1f1	1.02	1.43	0.7	1
61	2e12-	1.56	1.34	0.48	1.003
61	2f2	1.5	1.3	0.32	1.01
61	6b4	1.2	1.26	0.88	0.98
61	4c10	1.46	1.12	1.06	0.99
61	4g11	1.31	1.03	1.43	1.03
61	2f1	1.41	1.02	0.79	0.995
61	2g1	1.3	1	1.17	1

Experiment	Clone ID	Li	tau	Km	S
65	6g12	0.87	2.3	0.73	0.9605
65	1h6	0.84	2.2	1.62	0.9598
65	7f5	1.2	1.56	2.07	1.0087
65	5g5	2.3	1.49	0.45	0.9985
65	7h2	1.56	1.27	0.91	1.0658
65	7b2	1.98	1.16	0.6	0.9289
65	0g9	1.36	1.09	1.46	0.9927
65	6c7	1.48	1.06	0.86	0.9967
65	1e12-	1.59	1.05	1.03	0.9582
65	4e2-	1.21	1.05	1.11	0.943
65	6a10	1.7	1.04	0.93	0.992
65	4b9	1.48	1.04	1.61	1.0009
65	6c1	1.36	1.02	0.72	0.9978

Experiment	Clone ID	Li	tau	Km	S
68	2g6	1.39	3.9	1.17	0.9955
68	4g3	2	2.5	0.27	0.9927
68	5a3	1.04	1.64	0.65	0.8984
68	2b7	1.04	1.64	5.2	0.9237
68	5d10	2.75	1.36	0.73	1.0078
68	7d12	1.85	1.32	0.66	1.0084
68	7b9	1.8	1.19	0.56	1.0052
68	7b3	1.2	1.16	0.55	0.9951
68	1g10	1.48	1.05	1.22	1.0025

Experiment	Clone	Li	tau	Km	S
------------	-------	----	-----	----	---

ID					
70	2a7	1.94	4.6	0.7	1.0015
70	3d6	3.5	4.2	0.18	1.03
70	4f8	1.87	4.2	0.69	0.9979
70	7h5	2.4	2.6	0.18	1
70	5h6	3.1	2.3	0.6	0.999
70	7d6	3	2.2	2.29	0.9989
70	5a3	3.1	1.5	0.18	1.0058
70	7d2	2.5	1.4	0.66	1.0126
70	3h7	3.2	1.22	0.23	1.002
70	0h5	2.5	1.15	0.36	0.9992
70	0d7	1.86	1	1.83	0.993
70	1g12	2.42	1	0.26	0.965
Experiment	Clone ID	Li	tau	Km	S

71	1d10	1.6	4.5	1.06	1.0065
71	6f11	1.8	4.3	0.98	0.953
71	7h4	3.4	3.6	0.56	1.0045
71	4h3	3.1	3.1	0.42	1.0171
71	1h5	1.31	3.01	1.31	0.9421
71	5e4-	5.4	2.3	0.35	0.994
71	5c1	2.2	2.3	0.89	0.9746
71	0h7	3.6	1.8	0.59	1.0197
71	6h9	23.7	1.71	0.91	1.0064
71	7e3-	5.3	1.7	0.7	1.0028
71	5d4	11.1	1.48	0.35	1.0213
71	2e3-	4	1.47	0.45	0.9654
71	6h11	17.7	1.15	2.8	1.0064
71	2e10-	3	1.1	0.66	0.9588
71	2g2	4.4	1.01	0.44	1.0046

Control is  
PPE-2 71-  
5D4 at 60C

Experiment	Clone ID	Li	tau	Km	S
72	2g6	0.38	3.1	1.58	1.0052
72	5f12	0.81	1.53	1.02	0.9678
72	0d7	0.76	1.44	1.4	0.9838
72	5c12	0.87	1.43	1.04	0.9718
72	1e1-	1.04	1.41	1.15	0.9956
72	5b12	0.83	1.41	1.02	0.9731



72	0b7	1.11	1.04	0.91	1.0049
72	3b4	0.49	1.03	2.2	0.9581

Experiment Clone ID Li tau Km S

73	2h8	0.85	1.9	1.08	1.0123
73	4e6-	0.95	1.76	0.94	0.9939
73	3g8	0.86	1.53	1.04	1
73	1g3	1.7	1.14	0.97	0.9921

Experiment Clone ID Li tau Km S

74	2a9	0.96	1.77	0.86	0.999
74	4e10-	0.8	1.36	1.33	0.0989
					7
74	0d5	1.69	1.28	0.61	0.9927
74	6g7	1.75	1.07	1.33	1.0022
74	5d8	0.46	1.06	0.95	0.899
74	5e7-	1.22	1.05	0.87	0.9977
74	6e1-	1.19	1.02	0.96	0.999

Experiment Clone ID Li tau Km S

76	6c3	2.3	6.4	1.2	0.9865
76	2a9	0.93	4.7	1.08	0.999
76	3h9	1.26	2.6	1.02	0.9973
76	0b10	1.52	2.4	1.4	0.992
76	0h9	1.71	1.44	1.05	1.018
76	2e9-	0.44	1.15	1.2	0.9318
76	0e10-	1.67	1.1	1.02	1.014
76	0c10	1.13	1.05	1	0.9974
76	3e8-	1.35	1.03	1.1	0.9894
76	0d12	0.69	1	0.92	0.932
76	0f10	0.62	1	1.2	0.9478

Experiment Clone ID Li tau Km S

78	1e1-	0.54	8.9	1.15	0.9877
78	0h7	1.4	5	0.97	1.014
78	0a6	1	4.3	1.5	0.9967
78	0b10	1.93	2	1	0.9926
78	0f11	1.6	2	0.91	0.9905
78	3f1	2.4	1.7	1.09	0.9936
78	2b4	1.97	1.36	0.98	1.0094
78	5b3	3.2	1.19	1.03	0.9735

78	2g12	2.5	1.03	1	1.0134
78	0h2	1.6	1	1.15	1.0168

Control is  
PPE-2 78-  
0B10 at 62C

Experiment	Clone ID	Li	tau	Km	S
82	2g12	0.9811	2.09	0.8851	0.9939
82	4b9	1.0845	1.8419	0.8439	1.0078
82	0d1	0.7622	1.5171	1.11	0.9998
82	3g1	0.8805	1.504	0.9629	0.9927
82	1d1	0.9741	1.4497	0.8936	0.9986
82	1e8-	0.8206	1.4433	0.9876	0.9968
82	0h9	1.1355	1.3626	0.9171	1.0094
82	2c6	1.0931	1.3402	0.9482	1.0022
82	3g9	1.0364	1.251	0.968	1.0009
82	4h8	0.8816	1.1667	0.9165	1.0045
82	0a10	1.0535	1.1128	1.0413	1
82	4g1	1.4305	1.0862	1.1734	1.0059

Experiment	Clone ID	Li	tau	Km	S
84(121)	6h7	0.3755	29.363 9	2.3636	0.8905
84(121)	2h9	0.4264	28.795 8	1.819	0.904
84(121)	3f7	0.4161	25.305 8	1.8079	0.8988
84(121)	2h10	0.9667	14.465 8	0.8073	0.9947
84(121)	3a2	0.3329	12.6	2.5444	0.855
84(121)	3a6	1.2299	7.2384	0.7866	1.0046
84(121)	5b12	1.0535	6.0315	0.7824	1.0056
84(121)	5a7	1.0413	4.9054	0.8864	1.0071
84(121)	3d2	0.2032	4.8	2.4623	0.7973
84(121)	2a9	1.0847	4.7486	0.7746	1.0051
84(121)	5e11-	1.1918	4.0988	0.872	1.008
84(121)	7h2	0.9115	3.9929	0.909	1.0077
84(121)	3b5	1.2014	3.8251	0.7509	1.0086
84(121)	1f8	1.07	3.06	0.8276	1.0093
84(121)	2e2-	1.4356	1.9315	0.7863	1.0175

Control is  
PPE-2 84-

## 3a6 at 64C

Experiment	Clone ID	Li	tau	Km	S
85(86)	2a2	0.2266	12.901 3	3.326	0.8705
<b>85(86)</b>	<b>4f12</b>	<b>1.1167</b>	<b>4.7851</b>	<b>0.7439</b>	<b>1.0092</b>
85(86)	4e9-	1.0869	4.4953	0.8539	1.0068
85(86)	1f11	0.6994	4.0976	0.842	1.0124
85(86)	5a4	1.2273	4.09	0.9683	1.0098
85(86)	3e10-	0.8902	3.5342	0.8106	1.0069
85(86)	3e12-	1.0512	3.4883	0.853	1.0054
85(86)	5e4-	0.9562	3.3886	1.0328	1.0069
85(86)	0e6-	0.1494	3.0145	3.6293	0.8269
85(86)	6b1	0.7615	2.5712	0.8695	1.0055
85(86)	6h7	1.0285	2.5401	0.8963	1.0057
85(86)	4b11	0.9816	2.3899	0.7927	1.0063
85(86)	6d7	1.1087	2.0607	0.9042	1.0088
85(86)	2e10-	0.3028	2.0603	1.9649	0.8738
85(86)	2a9	1.448	1.1819	0.9722	1.0046

Control is  
PPE-2 85-  
4f12 at 65C

Experiment	Clone ID	Li	tau	Km	S
88	3c1	1.4439	2.0938	0.9874	0.9976
88	6g1	1.0184	1.2665	1.2184	1.0019
88	3e4-	1.331	1.0996	1.0669	0.9983

Experiment	Clone ID	Li	tau	Km	S
89	1a4	1.2565	2.4796	1.0338	0.997
89	3b1	0.7337	1.9976	0.9628	1.0001
89	2b12	1.0505	1.8496	1.0069	1.0012
89	0b5	1.5671	1.1362	1.0912	0.9995
89	1f1	1.378	1.1018	0.9804	0.996
89	2f1	1.4637	1.0894	0.9189	0.9992

Experiment	Clone ID	Li	tau	Km	S
90	0f1	1.4081	1.3632	1.027	0.9987
<b>90</b>	<b>1b5</b>	<b>1.4743</b>	<b>1.1154</b>	<b>1.0812</b>	<b>1.0011</b>
90	6g5	1.2756	1.0605	1.0462	1.0012
90	5e6-	1.0556	1.0569	1.1037	1.0011

90	4e3-	1.2934	1.0291	1.0733	1.0002

**TABLE 2: Stability Of Luciferase Activity At Different Temperatures (Half-Life In Hours)**

	<i>Room Temperature</i>	<i>37°C</i>	<i>50°C</i>	<i>60°</i>
Luc[T249M]	110	0.59	0.01	
49-7C6	430	68	31	6.3
78-0B10	3000	220	47	15

**TABLE 3: Michaelis-Menten Constants for Mutants Created by Directed Evolution**

	$K_m$ -luciferin	$K_m$ -ATP
Luc[T24]	0.32 $\mu$ M	18 $\mu$ M
49-7C6	0.99 $\mu$ M	14 $\mu$ M
78-0B10	1.6 $\mu$ M	3.4 $\mu$ M
90-1B5	2.2 $\mu$ M	3.0 $\mu$ M

**TABLE 4:**

Components	Concentration	Amount in 50 $\mu$	Final concentration
DATP	10 mM	1	0.2mM
DCTP	10 mM	1	0.2 mM
DGTP	10 mM	1	0.2 mM
DTTP	10 mM	1	0.2 mM
+pRAM18up	20 pmol/ $\mu$ l	1	0.4 pmol/ $\mu$ l
+pRAM19dn	20 pmol/ $\mu$ l	1	0.4 pmol/ $\mu$ l
PFU	2 U/ $\mu$ l	1	0.04 u/ $\mu$ L
°10x buffer	10x	5	1x
DNA		10 from purified wiz.	
Water		24.6	

**TABLE 5: Summary of Evolutionary Progression**

	①	Start with <i>LucPpe2</i> [T249M]
	②	Mutate 3 amino acids at N- and C-termini
5	③	Mutate 7 cysteines
	④	Perform two iterations of evolution → <i>Luc49-7C6</i>
	⑤	Mutagenesis of altered codons (9)
	⑥	Two iterations of evolution → <i>Luc78-0B10</i>
	⑦	Mutagenesis of consensus codons (28)
10	⑧	Mutagenesis of codon usage (24) → <i>Luc90-1B5</i>



**TABLE 6: One Iteration of Recursive Process**

5	①	1 clone → 3 libraries using error-prone PCR
		• 3 x Visual screen (~10,000 clones each)
		• 3 x Quantitative screen (704) clones each)
10	②	3 x 18 clones → library using sPCR
		• Visual screen (~10,000 clones)
		• Quantitative screen (704 clones)
	③	18 + 18 → library using sPCR
		• Visual screen (~10,000 clones)
		• Quantitative screen (704 clones)
	④	Output: 18 clones

## DOCUMENTS CITED

- Bowie, J.U. Reindhaar-Olsen, J.F. Lim, W.A., and Sauer, R.T., *Science*, (1990) 247:1306-1310.
- Fromant M, Blanquet S, Plateau P., *Analytical-Biochemistry* (1995) 224:347-53.
- 5 Hanahan, D (1985) in *DNA Cloning*, v. 1, ed. Glover, D.W. (IRL Press, Oxford) pp. 109-135.
- Hastings JW, Kricka LJ, Stanley PE, editors *Bioluminescence and Chemiluminescence, Molecular reporting with Photons*. Chichester: John Wiley & Sons (1996) 248-52.
- 10 Kajiyama N, Nakano E., *Biochemistry* (1993) 32:13795-13799.
- Kajiyama N, Nakano, E. *Biosci. Biotech.* (1994) 58(6):1170-1171.
- Leach, *et al.*, *Biochimica, et Biophysica Acta*, (1997) 1339(1):39-52.
- Stemmer, WP., *Proc. Natl. Acad. Sci. U.S.A.* (1994) 91:10747-51.
- 15 Stemmer U.S. Pat. No. 5,605,793.
- White, PJ *et al.*, editors *Bioluminescence and Chemiluminescence, Fundamentals and Applied Aspects*. Chichester: John Wiley & Sons (1994) 419-422.
- Wood KV, DeLuca M., *Analytical Biochemistry* (1987) 161:501-7.
- 20 Ye L, Buck LM, Schaeffer HJ, Leach FR., *Biochimica et Biophysica Acta* (1997) 1339:39-52.
- Stemmer, WP, *DNA* (1994) 91:10747-51.
- Fromant M, Blanquet S, Plateau P, *Analytical-Biochemistry* (1995) 224:347-53.
- 25 Saiki RK, Gefand DH, Stoffel S, Scharf SJ, Higuchi R, Mullis KB, Erlich HA, *Science* (1988) 239:487-91.

## WE CLAIM:

- 5           1.       A second beetle luciferase with increased thermostability as compared with a first luciferase, said second luciferase made by the following method:
- a)       mutating a polynucleotide sequence encoding the first luciferase to obtain a polynucleotide sequence encoding the second luciferase;
- b)       selecting the second luciferase if a plurality of characteristics including thermostability of a luciferase is in a preferred range.
- 10           2.       The second luciferase of claim 1, wherein the polynucleotide sequence encoding the first luciferase is the same as the sequence of Luc (T249M).
3.       The second luciferase of claim 1, wherein thermostability is at least 2 hours at about 50°C in aqueous solution.
- 15           4.       The second luciferase of claim 3, wherein thermostability is at least 5 hours at 50°C in aqueous solution.
5.       The second luciferase of claim 1, wherein the plurality of characteristics comprises brightness of luminescence, substrate utilization and luminescence signal.
- 20           6.       The second luciferase of claim 1, wherein the mutating is by directed evolution.
7.       A beetle luciferase that is thermostable for at least 2 hours at 50°C in aqueous solution.
- 25           8.       The luciferase of claim 7, that is thermostable for at least 5 hours at 50°C.

9. The luciferase of claim 7, wherein less than 5% luminescence activity is lost after incubation in solution for 2 hours at about 50°C.

10. A method for preparing a beetle luciferase with increased thermostability, said method comprising the following steps:

5 a) mutating a polynucleotide sequence encoding a first luciferase to obtain a sequence encoding a second luciferase; and

b) selecting the second luciferase if a plurality of characteristics including thermostability of a luciferase are in a preferred range.

10 11. The method of claim 10, wherein thermostability is at least 2 hours at 50°C.

12. The method of claim 11, wherein the thermostability is at least 5 hours at 50°C.

13. The method of claim 10, wherein mutating occurs at at least one position wherein a consensus amino acid is present in beetle species.

15 14. The method of claim 10, wherein mutating occurs at at least one position where a mutation occurred to produce the luciferase gene designated *luc90-1B5*.

20 15. A DNA molecule having a nucleotide sequence that encodes a mutant luciferase with increased thermostability as compared to the thermostability of a native luciferase.

16. The DNA molecule of claim 15, wherein the nucleotide sequence is selected from the group consisting of sequences.

a)

GGATCCAATGGAAGATAAAAAATATTTATATGGACCTGAACCATTTATCCCTTGGCTGA  
 TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTTACGGATG  
 CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAT  
 GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAACAAAACGACACAATAGC  
 GGTGTGTAGCGAAAATGGTTTGAATTTTTCTTCCATAATTGCATCATTGTATCTTGG  
 AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
 TATTGTAAACCACGCATAATTTTTTGTCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
 AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
 AGGTTATCAATGCCCAACACTTTTATTCTCAAAATCCGATATTAATCTTGACGTAAA  
 AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
 TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
 TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAAC  
 GGTATACCTTTCCACCATGGTTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
 ATTCAGGATTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGA  
 TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGTCAAAAAGTGC  
 ATTAGTTGAAAAGTACGATTATCGCACTTAAAGAAAATTCATCTGGTGGCGCACCTTT  
 ATCAAAAGAAATTTGGGGAGATGGTGAAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
 GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACAATGACGTCAGACC  
 GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCTACAACAGG  
 AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
 AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCG  
 CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
 GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACCTTT  
 ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
 GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
 AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTTT  
 GGATGAAATTTCCCAAGGATCAACTGGAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
 TGAAAAACACACCAATGGG\*

b)

GGATCCAATGGAAGATAAAAAATATTTATATGGACCTGAACCATTTATCCCTTGGCTGA  
 TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTTACGGATG  
 CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTGTTAAAT  
 GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAACAAAACGACACAATAGC  
 GGTGTGTAGCGAAAATGGTTTGAATTTTTCTTCCATAATTGCATCATTGTATCTTGG  
 AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
 TATTGTAAACCAACGACATAATTTTTTGTCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
 AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
 AGGTTATCAATGCCTCAACACTTTTATTCTCAAAATCCGATATTAATCTTGACGTAAA  
 AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
 TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
 TTCTCATGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAAC  
 GGTAAACCTTTCCACCATGGTTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
 ATTCAGGATTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGA  
 TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTTTTGTCAAAAAGTGC  
 ATTAGTTGAAAAGTACGATTATCGCACTTAAAGAAAATTCATCTGGTGGCGCACCTTT  
 ATCAAAAGAAATTTGGGGAGATGGTGAAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
 GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACAATGACGTCAGACC  
 GGGATCAAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCTACAACAGG  
 AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
 AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCG  
 CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
 GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACCTTT  
 ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
 GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
 AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTTT  
 GGATGAAATTTCCCAAGGATCAACTGGAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
 TGAAAAACACACCAATGGG\*

c)

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
 TGGGACGGCTGGGAGAACAGATGTTTACGCATTATCTCGTTATGCAGATATTTTCAGGATG  
 CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
 GTCGTGTCGTTTAGCGGAAAGTTTTTAAAGATATGGATTAAACAAAACGACACAATAGC  
 GGTGTGTAGCGAAAATGGTTTGCAATTTTCCCTTCTATAATTGCATCATTGTATCTTGG  
 AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
 TATTGTAAACCACGCATAATTTTTTGTCCAAAGAATACTTTTCAAAAAGTACTGAATGT  
 AAAATCTAAATTAATAATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
 AGGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAA  
 AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
 TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGTACGATT  
 TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAC  
 GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
 ATTCGGAGTTGTTCTAATGCACACGTTTGAAGAAAACCTATTCTACAATCATTACAAGA  
 TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGTCAAAAAGTGC  
 ATTAGTTGAAAAGTACGATTATCGCACTTAAAGAAATGTCATCTGGTGGCGCACCTTT  
 ATCAAAAGAAATGGGGAGATGGTGAAGAAACGGTTTAAATTAACCTTTGTGAGGCAAGG  
 GTATGGATTAACGAAACCACTTCGGCTGTTTAAATACACCGAACCAATGACGTGAGACC  
 GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCCTACAACAGG  
 AAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGGCGACATGATAATGAA  
 AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTACCAAGACGGATGGTTGCG  
 CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
 GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
 ACAACATCCGTATATTGTTGATGCCGCGCTTACTGGTATACCGGATGAAGCCGCGGGCGA  
 GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
 AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
 GGATGAAATCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
 TGAAAAACACACCAATGGG\*

d)

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
 TGGGACGGCTGGGAGAACAGATGTTTACGCATTATCTCGTTATGCAGATATTTTCAGGATG  
 CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
 GTCGTGTCGTTTAGCGGAAAGTTTTTAAAGATATGGATTAAACAAAACGACACAATAGC  
 GGTGTGTAGCGAAAATGGTTTGCAATTTTCCCTTCTATAATTGCATCATTGTATCTTGG  
 AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
 TATTGTAAACCACGCATAATTTTTTGTCCAAAGAATACTTTTCAAAAAGTACTGAATGT  
 AAAATCTAAATTAATAATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
 AGGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAA  
 AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
 TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
 TTCTATTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAC  
 GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
 ATTCGGATGTTTCTAATGCACACGTTTGAAGAAAACCTATTCTACAATCATTACAAGA  
 TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGTCAAAAAGTGC  
 ATTAGTTGAAAAGTACGATTATCGCACTTAAAGAAATGTCATCTGGTGGCGCACCTTT  
 ATCAAAAGAAATGGGGAGATGGTGAAGAAACGGTTTAAATTAACCTTTGTGAGGCAAGG  
 GTATGGATTAACAGAAACCACTTCGGCTGTTTAAATACACCGAACCAATGACGTGAGACC  
 GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCCTACAACAGG  
 AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGGCGACATGATAATGAA  
 AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTAAACAAAGACGGATGGTTGCG  
 CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
 GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
 ACAACATCCGTATATTGTTGATGCCGCGCTTACTGGTATACCGGATGAAGCCGCGGGCGA  
 GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
 AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
 GGATGAAATCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
 TGAAAAACACACCAATGGG\*

e)

GGATCCAATGGAAGATAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTTTCAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTTAAAGATATGGATTAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAATGGTTTGCAATTTTCTTCTTATAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCGCATAATTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATTTTAAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCATGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTGAAGAAAACTATTCTACAATCATTACAAGA  
TTATAAGTGAAAGTACTTTACTTGTACCAACATTAATGGCATTTTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAAATGCAATCTGGTGGCGCACCTTT  
ATCAAAAGAAATGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTAAATTACACCGAACAATGACGTCAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGGCGACATGATAATGAA  
AGGTATTATAATAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAAATATAAAGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATCCCAAAGGATCAACTGGAAAATGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACACCAATGGG\*

f)

GGATCCAATGGCAGATAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTTTCAAGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTTCCTTCCGTAATGCATCATTGTATCTTGGA  
ATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGGT  
ATTGTAAAACACGCATAATTTTGTCTCCAAGAATACTTTTCAAAAAGTACTGAATGTA  
AAATCTAAATTAATCTGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGGA  
GGTTATCAATGCCTCAACAACCTTATTTCTCAAAATTCGGATATTAATCTTGACGTAAAA  
AAATTTAAACCATAATCTTTTAAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTCT  
GGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATTT  
TCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCCACGACAGCAATTTTAACG  
GTAATACCTTTCCACCATGGTTTTGGTATGATgACCACATTAGGATACTTTACTTGTGGA  
TTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGAT  
TATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGCA  
TTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATGTCATCTGGTGGCGCACCTTTA  
TCAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTTAAATTAACTTTGTCAGGCAAGGG  
TATGGATTAACAGAAACCCTTCGGCTGTTTTAATTACACCGAAAxxxxxxGCCAGACCG  
GGATCAACTGGTAAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGGA  
AAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGGCGCCATGATAATGAAG  
GGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCGC  
TCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAAG  
TCATTAATTAAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTTA  
CAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATTCGGATGAAGCCGCGGGCGAG  
CTTCCAGCTGCAGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACAA  
GATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTTG  
GATGAAATTCCTAAAGGATCAACTGAAAAATTGACAGAAAAGTGTAAAGACAAATGTTT  
GAAAAACACACCAATGGG\*



g)

GGATCCAATGGCAGATAAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTACGCATTATCTCGTTATGCAGATATTTAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTCCCTCCTGTAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAACCACGCATAATTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACTTTATTCTCAAAATTCGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATCTTTTAAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
TGGTACAACCTGGTGTCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAATTCATCTGGTGGCGCACCTTT  
ATCAAAAAGAAATGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAA~~xxxxxx~~GTCAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTGGGGCCAAATGAACCTGGAGAATTGTTTTAAAGGCGACATGATAATGAA  
AGGTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTGTTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGGGTGGGGTGAATTTTT  
GGATGAAATTCCTAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACACCAATGGG\*

h)

GGATCCAATGGCAGATAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTCCCGGATG  
CATAGCATTGCACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCATATTTCCCTCCTGTAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCACGCATAATTTTTTGTCTCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAAATATGTAGAAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACTTTATTCTCAAAATCCGATATTAACTTTGACGTAAA  
AAAATTAAACCAAATCTTTTAAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCCGAAGGGAGTCATGCTAACTCACAGAATATTGTTGCACGATT  
TTCTATTGCAAAAGATCCTACTTTTGGTAACGCAATTAAATCCAACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTCTAATGCACACGTTTGAAGAAAACATTCTTACAATCATTACAAGA  
TTATAAAGTGGAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTAAATTAAACTTTGTGAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTTAATTACACCGAAAAGGCGCAGACC  
GGGATCAACTGGTAAAATAGTACCATTTACGCTGTAAAGTTGTGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGGCGCCATGATAATGAA  
GGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACACCAATGGG\*

i)

GGATCCAATGGCAGATAAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
 TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTCCCGGATG  
 CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
 GTCGTGTCGTTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
 GGTGTGTAGCGAAAATGGTTGCAATTTTTCTTCTCTGTAATTGCATCATTGTATCTTGG  
 AATAATTGCAGCACCTGTAGTGATAAATACGTTGAACGTGAATTAATACACAGTCTTGG  
 TATTGTAAAACCACGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
 AAAATCTAAATTAATAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
 AGGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATTCGATAGTAATCTGGACGTAAA  
 AAAATTTAAACCAATTCCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
 TGGTACAACCTGGTGTCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
 TTCTCTTGCAAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTAAC  
 GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
 ATTCGGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
 TTATAAAGTGGAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
 ATTAGTTGAAAAGTACGATTATTCGCACTTAAAAGAAATTGCATCTGGTGGCGCACCTTT  
 ATCAAAAGAAATTGGGGAGATGGTGA AAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
 GTATGGATTAAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAAxxxxxxGCCAGACC  
 GGGATCAACTGGTAAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGG  
 AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGCGCCATGATAATGAA  
 GGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
 CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
 GTCATTAAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
 ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
 GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
 AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTT  
 GGATGAAATCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
 TGAAAAACACACCAATGGG\*

7)

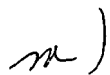
GGATCCAATGGCAGATAAAAAATATTTTATATGGGCCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTCGGGGCTG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAAAT  
GTCGTGTCGTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAATGGTTTGCATTTTTTCCTTCTGTAAATGCATCATTGTATCTTGG  
AATAATTGTGGCACCTGTTAACGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAGTTTTTGTCTCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAACTTAAATTAATACTGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACTTTATTCTCAAAATTCGGATATTAATCTTGACGTAAA  
AAAAATTAACCATATCTTTTAAATCGAGACGATCAGGTTGCGTTGATTATGTTTTCTTC  
TGGTACAACGGTCTGCCGAAGGGAGTCATGCTAACTCACAGAATATTGTTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCCACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTAAATTAACTTTGTCAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTTAATTACACCGAAAAGGCGCAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGGCCCGATGATAATGAA  
GGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATTCCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAATAATCTAAACGAACAAATCGTACA  
AGATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTTT  
GGATGAAATTCCTCAAGGATCAACTGGAATAATGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACCAATGGG

h)

GGATCCAATGGCAGATAAGAATATTTTATATGGGCCCGAACCATTTATCCCTTGGAAGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATCCGGGCTG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTCTGAAACT  
GTCGTGTCGTTTAGCGGAAAGTTTAAAAAGTATGGATTAAAAACAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTCTGCAATTTTCTCTTCTGTAATTGCATCATGTATCTTGG  
AATAATTGTGGCACCTGTTAACGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAATTTTTTGTCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAACTAAATTAATCTGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCCTCAACAACTTTATTTCTCAAAATTCGGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGTTAATGTTTTCTTC  
TGGTACAACTGGTCTGCCGAAGGGAGTCATGCTAACTCACAAGAAATATTGTTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCCACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGTCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAAGAAATTGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTTAATTACACCGAAA~~xxxxxx~~GCCAAACC  
GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTGATCCTACAACAGG  
AAAAATTTGGGGCCAAATGAACCTGGAGAATTGTATTTAAAGGCCCGATGATAATGAA  
GGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCACTGATTAAATATAAAGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATCCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AGATTATGTTGCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATTCCTAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACACCAATGGG

2)

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGATGGGACGGCTGGAGAACAG  
ATGTTTTACGCATTATCTCGTTATGCAGATATTTAGGATGCATAGCATTGACAAATGCTCATACAAAAGAAAATGTTT  
TATATGAAGAGTTTTTAAAAATTGTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAACAAAACGACACAAT  
AGCGGTGTGTAGCGAAAATGGTTTGAATTTTTCCTTCCCTTAATTGCATCATTGTATCTTGGAATAATTGCAGCACCT  
GTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGGTATTGTAAAACCACGCATAATTTTTTGTTCCAAGA  
ATACTTTTCAAAAAGTACTGAATGTAAATCTAAATTAATAATATGTAGAACTATTATTATATTAGACTTAAATGAAGA  
CTTAGGAGGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAAAAAATTTAAACCA  
AATTCTTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTCTGGTACAACCTGGTGTTCGAAGGGAGTCATGC  
TAACTCACAGAATATTGTTGCACGATTTTCTCATTTGCAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGC  
AATTTAACGGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGGATTCCGAGTT  
GCTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGATTATAAAGTGGAAAGTACTTTACTTGTAC  
CAACATTAATGGCATTTTTTGCAAAAAGTGCATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTGCATCTGG  
TGGCGCACCTTTATCAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTAAATTAAACTTTGTGAGGCAAGGGTATGGA  
TTAACAGAAACCACTTCGGCTGTTTTAATTACACCGGACACTGACGTCAGACCGGGATCAACTGGTAAATAGTACCAT  
TTCACGCTGTTAAAGTTGTCGATCCTACAACAGGAAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGG  
CGACATGATAATGAAAAGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCGCTCT  
GGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAAGTCATTAATTAAATATAAAGGTT  
ATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTTACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACC  
GGATGAAGCCGCGGGCGAGCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACAA  
AATTTGTTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTTTTGGATGAAATTCCCAAAGGAT  
CAACTGGAATAATTGACAGAAAAGTGTAAAGACAAATGTTTGAATAACACAAATCTAAGCTG



GGATCCCATGATGAAGCGAGAGAAAAATGTTATATATGGACCCGAACCCCTACACCCCTT  
GGAAGACTTAACAGCTGGAGAAATGCTCTTCCGTGCCCTTCGAAAACATTCTCATTTACC  
GCAGGCTTTAGTAGATGTGGTTGGCGACGAATCGCTTTCCTATAAAGAGTTTTTTGAAGC  
GACAGTCCTCCTAGCGCAAAGTCTCCACAATTGTGGATACAAGATGAATGATGTAGTGTC  
GATCTGCGCCGAGAATAATACAAGATTTTTTATTCCCGTTATTGCAGCTTGGTATATTGG  
TATGATTGTAGCACCTGTTAATGAAAGTTACATCCCAGATGAACTCTGTAAGGTGATGGG  
TATATCGAAACCACAAATAGTTTTTACGACAAAGAACATTTTAAATAAGGTATTGGAGGT  
ACAGAGCAGAACTAATTTTCATAAAAAGGATCATCATACTTGATACTGTAGAAAACATACA  
CGGTTGTGAAAGTCTTCCCAATTTTATTTCTCGTTATTCGGATGGAATATTGCCAACTT  
CAAACCTTTACATTTTCGATCCTGTTGAGCAAGTGGCAGCTATCTTATGTTTCGTGAGGCAC  
TACTGGATTACCGAAAGGTGTAATGCAAACTCACCAAAATATTTGTGTCCGACTTATACA  
TGCTTTAGACCCCGAGGACAGGAACGCAACTTATTCCTGGTGTGACAGTCTTAGTATATCT  
GCCTTTTTTCCATGCTTTTGGGTTCTCTATAACCTTGGGATACTTCATGGTGGGTCTTCG  
TGTTATCATGTTTCAGACGATTTGATCAAGAAGCATTTCTAAAAGCTATTCAGGATTATGA  
AGTTCGAAGTGAATTAACGTTCCATCAGTAATATTGTTCTTATCGAAAAGTCCTTTGGT  
TGACAAATACGATTTATCAAGTTTAAGGGAATTGTGTTGCGGTGCGGCACCATTAGCAAA  
AGAAGTTGCTGAGGTTGCAGCAAAACGATTAAACTTGCCAGGAATTCGCTGTGGATTGG  
TTTGACAGAATCTACTTCAGCTAATATACACAGTCTTAGGGATGAATTTAAATCAGGATC  
ACTTGGAAGAGTTACTCCTTTAATGGCAGCTAAAATAGCAGATAGGGAACTGGTAAAGC  
ATTGGGACCAAATCAAGTTGGTGAATTATGCATTAAAGGTCCCATGGTATCGAAAGGTTA  
CGTGAACAATGTAGAAGCTACCAAAGAAGCTATTGATGATGATGGTTGGCTTCACTCTGG  
AGACTTTGGATACTATGATGAGGATGAGCATTTCTATGTGGTGGACCGTTACAAGGAATT  
GATTAAATATAAGGGCTCTCAGGTAGCACCTGCAGAACTAGAAGAGATTTTATTGAAAAA  
TCCATGTATCAGAGATGTTGCTGTGGTTGGTATTCTGATCTAGAAGCTGGAGAAGTGGC  
ATCTGCGTTTGTGGTTAAACAGCCCGAAAGGAGATTACAGCTAAAGAAGTGTACGATTA  
TCTTGCCGAGAGGGTCTCCCATACAAAGTATTTGCGTGGAGGGGTTTCGATTCGTTGATAG  
CATACCAAGGAATGTTACAGGTAAAATTACAAGAAAGGAAGTCTGAAGCAGTTGCTGGA  
GAAGGCGGGAGGT

n) GGATCCCATGATGAAGCGAGAGAAAAATGTTATATATGGACCCGAACCCCTACACCCCTT  
GGAAGACTTAACAGCTGGAGAAATGCTCTTCCGTGCCCTTCGAAAACATTCTCATTTACC  
GCAGGCTTTAGTAGATGTGGTTGGCGACGAATCGCTTCTATAAAGAGTTTTTTGAAGC  
GACAGTCCCTCCTAGCGCAAAGTCTCCACAATTGTGGATACAAGATGAATGATGTAGTGTC  
GATCTGCGCCGAGAATAATACAAGATTTTTTATTCCCGTTATTGCAGCTTGGTATATTGG  
TATGATTGTAGCACCTGTTAATGAAAGTTACATCCCAGATGAACTCTGTAAGGTGATGGG  
TATATCGAAACCACAAATAGTTTTTACGACAAAGAACATTTTAAATAAGGTATTGGAGGT  
ACAGAGCAGAACTAATTTTATAAAAAGGATCATCATACTTGATACTGTAGAAAACATACA  
CGGTTGTGAAAGTCTTCCCAATTTTTATTCTCGTTATTCGGATGGAAATATTGCCAACTT  
CAAACCTTTACATTTCGATCCTGTTGAGCAAGTGGCAGCTATCTTATGTTTCGTCAAGGCAC  
TACTGGATTACCGAAAGGTGTAATGCAAACCTCACCAAATATTTGTGTCCGACTTATACA  
TGCTTTAGACCCCAGGGCAGGAACGCAACTTATTCTGGTGTGACAGTCTTAGTATATCT  
GCCTTTTTTCCATGCTTTTGGGTTCTCTATAACCTTGGGATACTTCATGGTGGGTCTTCG  
TGTATCATGTTTCAGACGATTTGATCAAGAAGCATTTCATAAAGCTATTACAGGATTATGA  
AGTTCGAAGTGTAATTAACGTTCCATCAGTAATATTGTTCTTATCGAAAAGTCCTTTGGT  
TGACAAATACGATTTATCAAGTTTAAAGGAATTGTGTTGCGGTGCGGCACCATTAGCAAA  
AGAAGTTGCTGAGGTTGCAGCAAAACGATTAAACTTGCCAGGAATTCGCTGTGGATTGG  
TTTGACAGAATCTACTTCAGCTAATATACACAGTCTTAGGGATGAATTTAAATCAGGATC  
ACTTGGGAAGAGTTACTCCTTTAATGGCAGCTAAAATAGCAGATAGGGAACTGGTAAAGC  
ATTGGGACCAAATCAAGTTGGTGAATTATGCATTAAAGGTCCCATGGTATCGAAAGGTTA  
CGTGAACAATGTAGAAGCTACCAAAGAAGCTATTGATGATGATGGTTGGCTTCACTCTGG  
AGACTTTGGATACTATGATGAGGATGAGCATTCTATGTGGTGGACCGTTACAAGGAATT  
GATTAATATAAGGGCTCTCAGGTAGCACCTGCAGAACTAGAAGAGATTTATTGAAAAA  
TCCATGTATCAGAGATGTTGCTGTGGTTGGTATTCCTGATCTAGAAGCTGGAGAAGTACC  
ATCTGCGTTTGTGGTTAAACAGCCCGAAAGGAGATTACAGCTAAAGAAGTGACGATTA  
TCTTGCCGAGAGGGTCTCCCATACAAAGTATTTGCGTGGAGGGGTTTCGATTCGTTGATAG  
CATACCAAGGAATGTTACAGGTAAATTACAAGAAAGGAAGTCTGAAGCAGTTGCTGGA  
GAAGGCGGGAGGT



Wood and Hall

17. A DNA molecule having a nucleotide sequence that encodes a luciferase of claim 1 or 7.

18. The use of luciferases of claims 1 or 7 in ATP assays; as luminescent labels for nucleic acids, proteins, or other macromolecules; as genetic reporters; in enzyme immobilization; as hybrid proteins; in high temperature reactors; and in luminescent solution.

19. A kit comprising a beetle luciferase with a half-life of at least 2 hours at 50°C.

20. The kit of claim 19 used for ATP assays; as luminescent labels for nucleic acids, proteins, or other macromolecules; as genetic reporters; in enzyme immobilization; as hybrid proteins; in high temperature reactors; and in luminescent solution.

21. A luciferase having an amino acid sequence consisting of

a)

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCLAESFKKYGLKQNDTIAVCSENGLQFFLPITIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKYVETIIILDLNEDLGGYQCLNMFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGFTCGFRVVLMTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKRFRKLNFRQGYGLTETTSAVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEEATKAI INKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVOTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG



DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEELLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPPIIASLYLGI AAPVSDKYIERELIHS LG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISONSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSHAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVMHTFEEKLFLQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTS AVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEEATKAI INKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYQYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

c)

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPPIIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVVRFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMITLGYFTCGFRVVLMTFEKLFQLSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKRFLNFVRQGYGLTETTSAVLITPNNVVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYYNNEEATKAIITKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPPIIASLYLGIIAAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSIAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVMHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTSAVLITPNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNEEATKAIINKOGWLR  
SGDIAYYDNDGHFYIVDRKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

*e)*

DPMEDKNILYGPEPFYPLADGTAGEQMFDAISRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPPIIASLYLGIIAAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISONSDINLDVK  
KFKPYSENRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSAKDPTEFGNAINPTTALT  
VIPFHHGFGMMTTLGYFTCGFRVLMHTFEEKLFLQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNEEATKAI INKOGWLR  
SGDIAYDNDGHEFYIVDRKSLIKYQYVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

A)

DPMADKNILYGPEPFYPLADGTAGEQMFDAISRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIASLYLGIIAAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKSVEIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRRDQVALVMFSSGTTGVSKGVMLTHKNIVARFSLAKDPTFGNAINPTTALT  
VIPFHHGFGMMTTLGYFTCGFRVLMHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPK~~CK~~ARPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGAMIMKGYYNNEEATKAIIDNDGWLRS  
GDIAYDNDGHFYIVDRKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVQTGKYLNEQIVQDFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
KHTNGS

g)

DPMAADKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVMHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPKKVRPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGDMIMKGYYNNEEATKAIIDKDWLRS  
GDIAYYDNDGHFYIVDRLKSLIKYGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG



h)

DPMADKNILYGPEPFYPLADGTAGEQMFDALSRADIPGCCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOQYFLPVIASLYLGI AAPVSDKYIERELIHSIG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNFIQNSDINLDVK  
KFKPNSFNRDDQVALVMFSSGTTGVPKGVMLTHKNIVARFSIAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPK~~Q~~ARPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGAMIMKGYNNNEEATKAIIDKDGWLRS  
GDIAYYDNDGHFYIVDRLKSLIKYKGQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

DPMAKKNILYGPEPFYPLADGTAGEQMFDALSRADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIAASLYLGI AAPVSDKYVERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNMFISQNSDSNLDVK  
KFKPNSFNRDDQVALVMFSSGTTGVKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMITLGYFTCGFRVVMHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTSAVLITPK~~Q~~ARPG  
STGKIVPFHAVKVVDPTTGKILGPNPGEYFKGAMIMKGYNNNEEATKAIIDKDGWLRS  
GDIAYYDNDGHFYIVDRLKSLIKYKGQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

7)

DPMADKNILYGPEPFYPLADGTAGEQMFDAISRYADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIASLYLGIIVAPVNDKYIERELIHSLG  
IVKPRIVFCSKNTFQKVLNVKSKLSVETIIILDNLNEDLGGYQCLNMFISQNSDINLDVK  
KFKPYSFNRDDQVALIMFSSGTTGLPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMTTLGYFTCGFRVVLNHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTSAVLITPK~~AR~~PG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGPMIMKGYNNNEEATKAIIDNDGWLRS  
GDIAYYDNDGHFYIVDRKSLIKYKGQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQDFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG



DPMADKNILYGPEPFYPLEDGTAGEQMFDALSRADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIASLYLGIIVAPVNDKYIERELIHS LG  
IVKPRIIFCSKNTFQKVLNVKSKLKS VETIIILDLNEDLGGYQCLNMFISQNSDINLDVK  
KFKPYSFNRDDQVALLMFSSGTTGLPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVMHTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKRKLNFVRQGYGLTETTS AVLITPK~~AK~~AKPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGPMIMKGYNNNEATKAIIDNDGWLRS  
GDIAYYDNDGHFYIVDRLSLIKYG YQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
\_AAGVVVQTGKYLNEQIVQDYVASQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG



DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPLIASLYLGIIAAPVSDKYIERELIHS LG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPNSFNRRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFESHCKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLG YFTCGFRVALMHTFEEKLFLQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTS AVLITPDTDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKSYNNNEEATKAIINKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYKG YQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHKSKL

DPMMKREKNVIYGPEPLHPLEDLTAGEMLFRALRKHSHLPQALVDVVGDESLSYKEFFEA  
TVLLAQSLHNCGYKMNDVVSICAENNTRFFIPVIAAWYIGMIVAPVNESYIPDELCKVMG  
ISKPQIVFTTKNII LNKVLEVQSRTNFIKRIIILDTVENIHGCESLPNFISRYSDGNIANF  
KPLHFDPEQVAAILCSSGTTGLPKGVMQTHQNICVRLIHALDPAGTQLIPGVTVLVYL  
PFFHAFGFSITLGYFMVGLRVIMFRREDQEAFLKAIQDYEVRSVINVPSVILFLSKSPLV  
DKYDLSSLRELCCGAAPLAKEVAEVAARKRLNLPGIRCGFGLTESTSANIHSRDEFKSGS  
LGRVTPLMAAKIADRETGKALGPNQVGELCIKGPMVSKGYVNNVEATKEAIDDDGWLHSG  
DFGYDEDEHFYVVDYKELIKYKGSQVAPAELEEILLKNPCIRDVAVVGIPDLEAGELP  
SAFVVKQPGKEITAKEVYDYLAERVSHTKYLGGVRFVDSIPRNVTKITRKELLKQLE  
KAGG

n) DPMMKREKNVIYGPEPLHPLEDLTAGEMLFRALRKHSHLPQALVDVVGDESLSYKEFFEA  
TVLLAQSLHNCGYKMNDVVSICAENNTREFFIPVIAAWYIGMIVAPVNESYIPDELCKVMG  
ISKPQIVFTTKNILNKVLEVQSRTNFIKRIIILDTVENIHGCESLPNFISRYSDGNIANF  
KPLHFDPEQVAAILCSSGTTGLPKGVMQTHQNICVRLIHALDPRAGTQLIPGVTVLVYL  
PFFHAFGFSITLGYFMVGLRVIMFRRFDQEAFLKAIQDYEVRSVINVPSVILFLSKSPLV  
DKYDLSSLRELCCGAAPLAKEVAEVAAKRLNLPGIRCGFGLTESTSANIHSLRDEFKSGS  
LGRVTPLMAAKIADRETGKALGPNQVGELCIKGPMVSKGYVNNVEATKEAIDDDGWLHSG  
DFGYDEDEHFYVVDYKELIKYKGSQVAPAELEEILLKNPCIRDVAVVGIPDLEAGELP  
SAFVVKQPGKEITAKEVYDYLAERVSHTKYLRGGVRFVDSIPRNVTKITRKELLKQLE  
KAGG

22. The luciferase of claim 21 further characterized as having a half-life of 2 hours at 50°C.



22. The luciferase of claim 21 further characterized as having a half-life of 2 hours at 50°C.

## Stability at 37C normalized to t=0

Fig 1

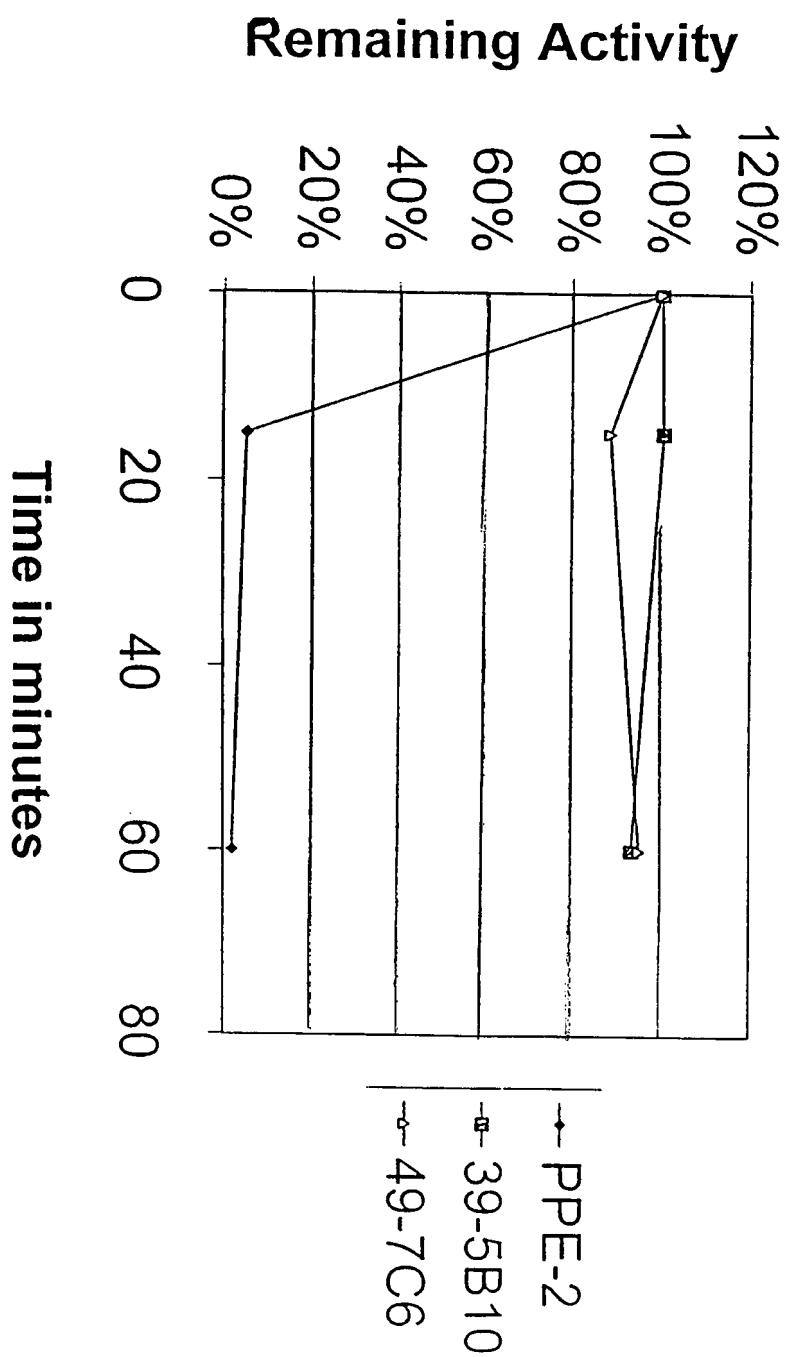
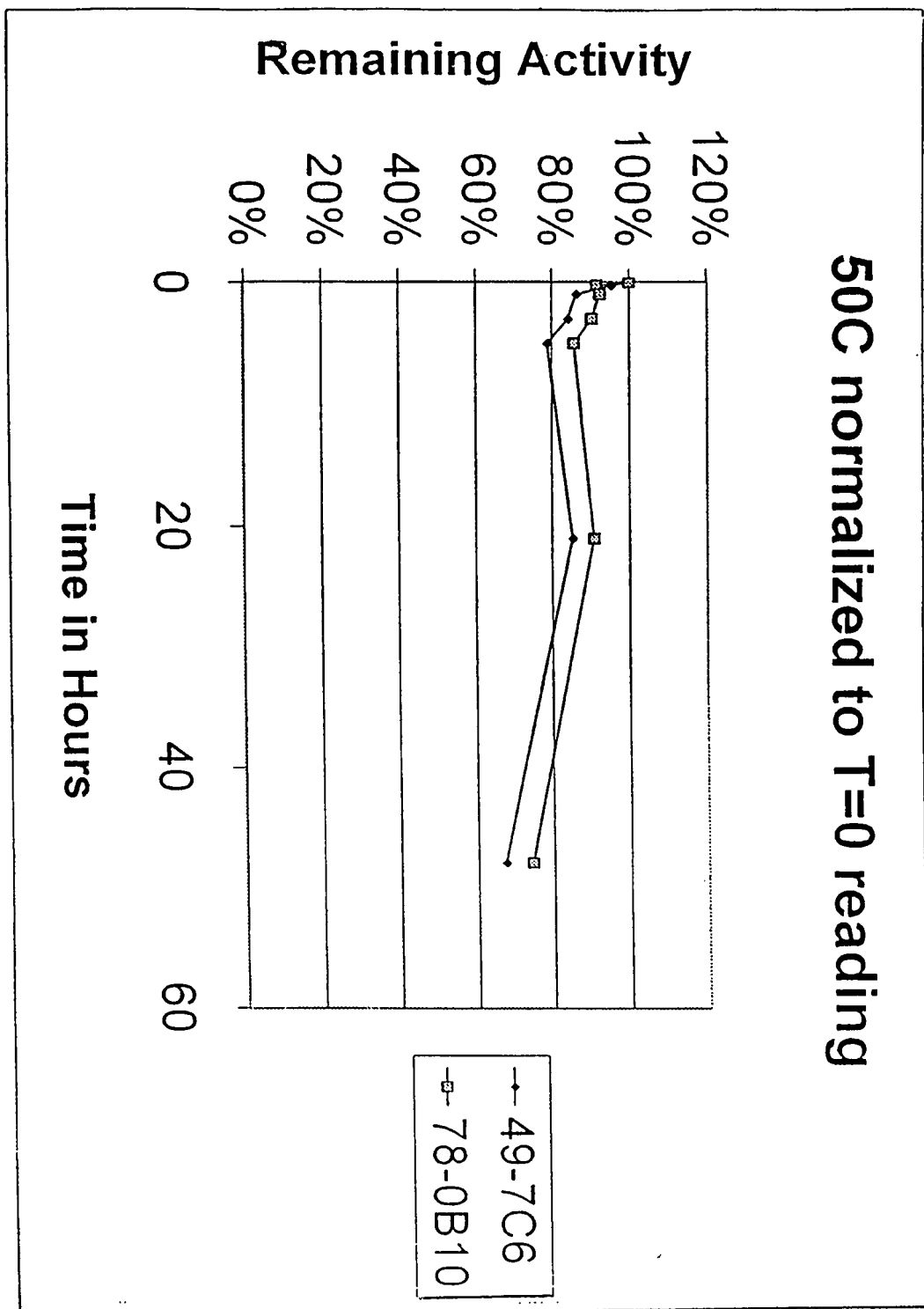


Fig. 2



3/59

Fig 3

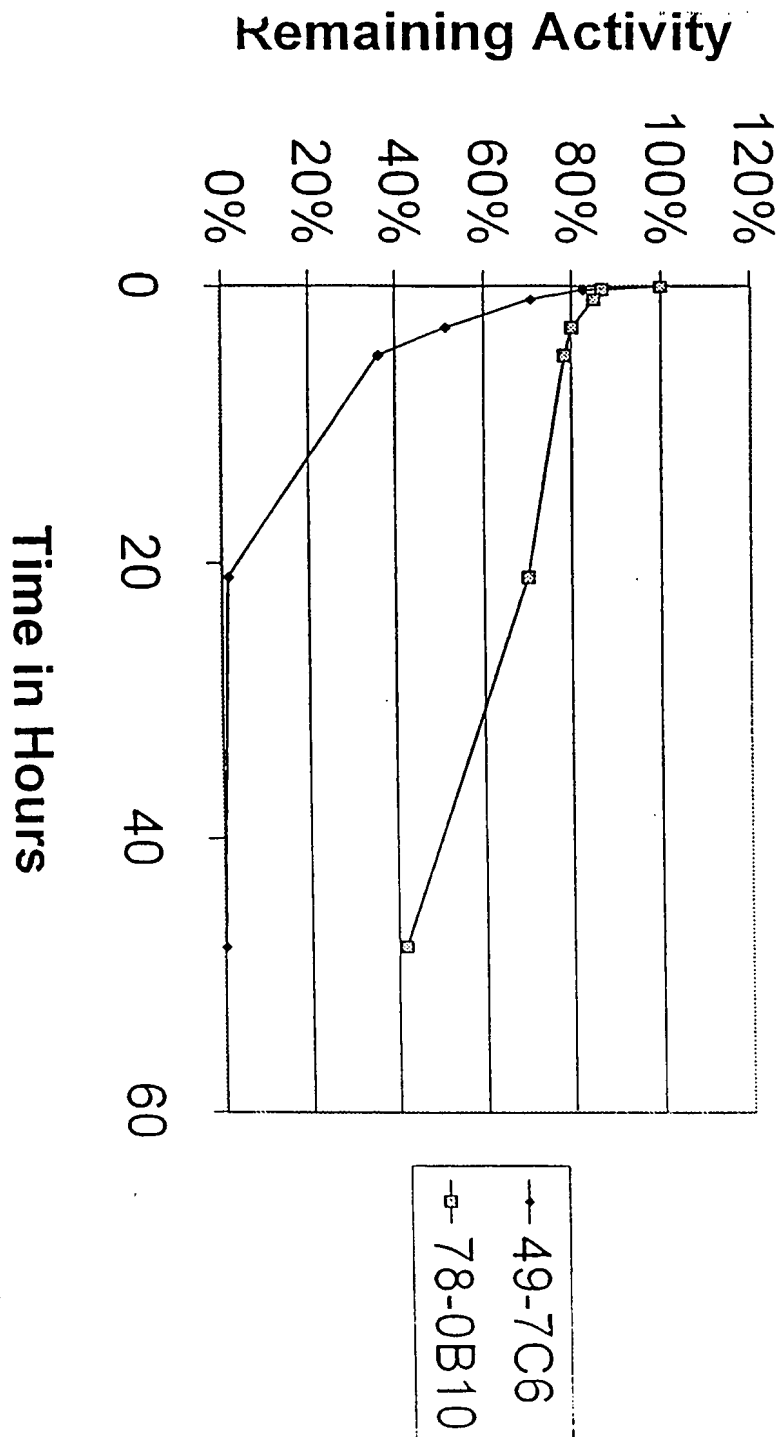
**60C Luminescence normalized to t=0**

Fig 4

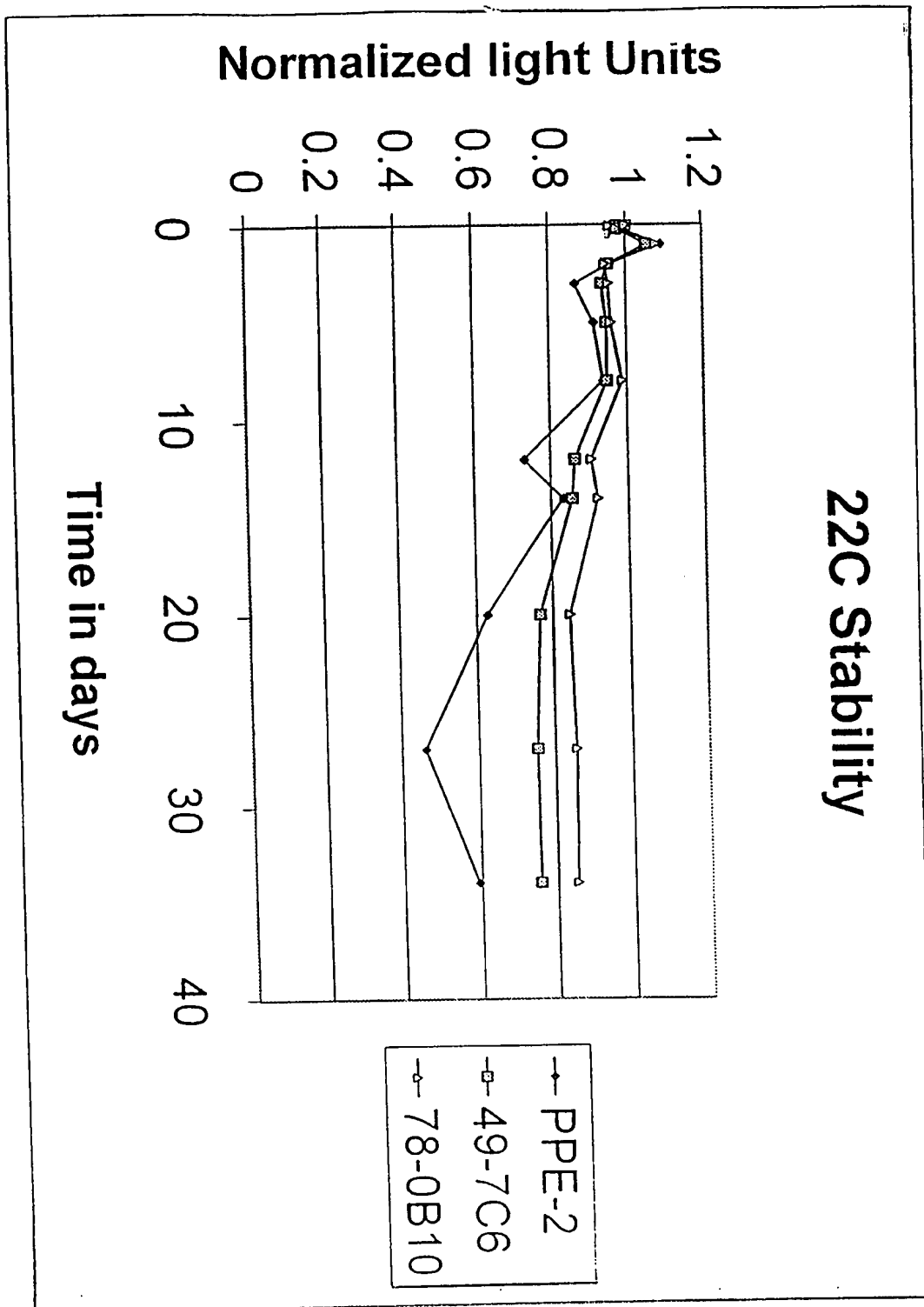


Fig 5

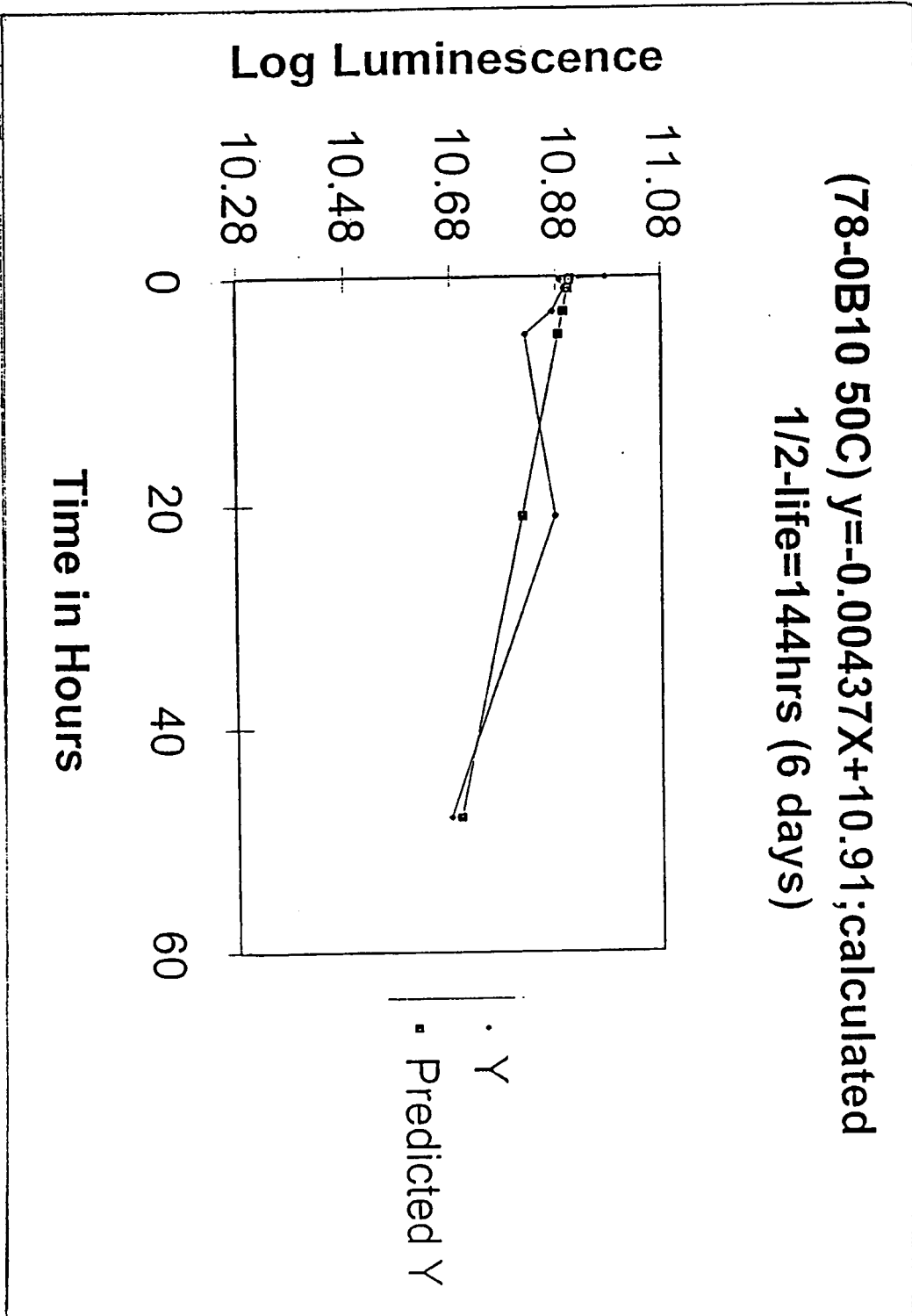
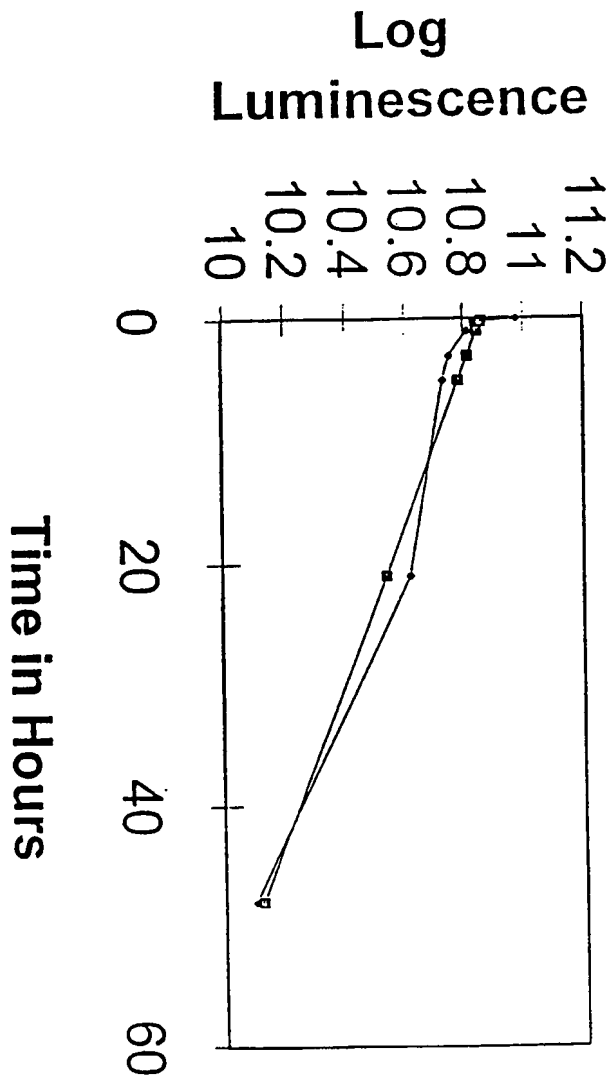


Fig 9

(78-0B10 60C)  $y = -0.154X + 10.86$ ; Calculated 1/2-life = 38 hours (1.58 days)



• Y  
■ Predicted Y

Fig 7

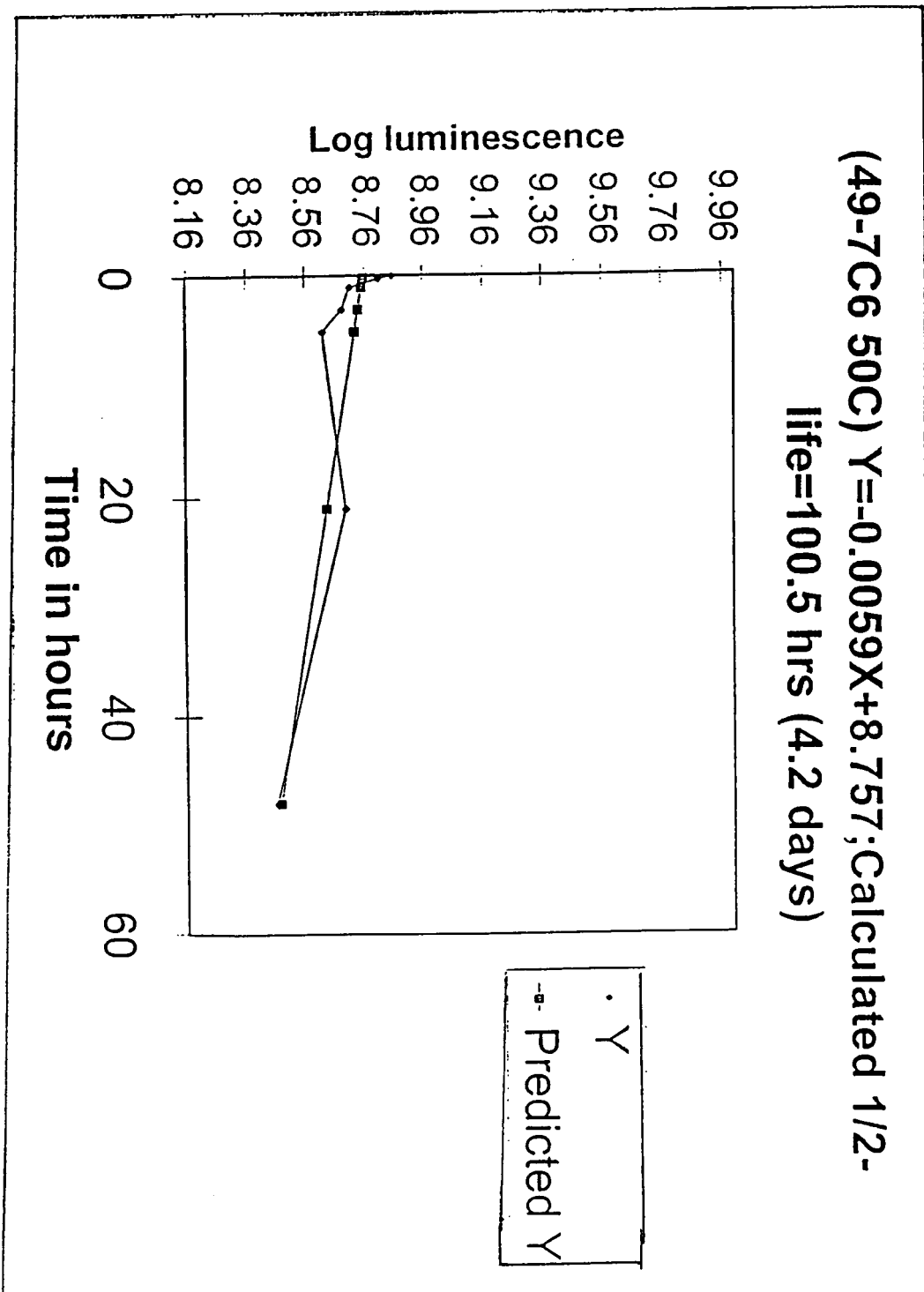
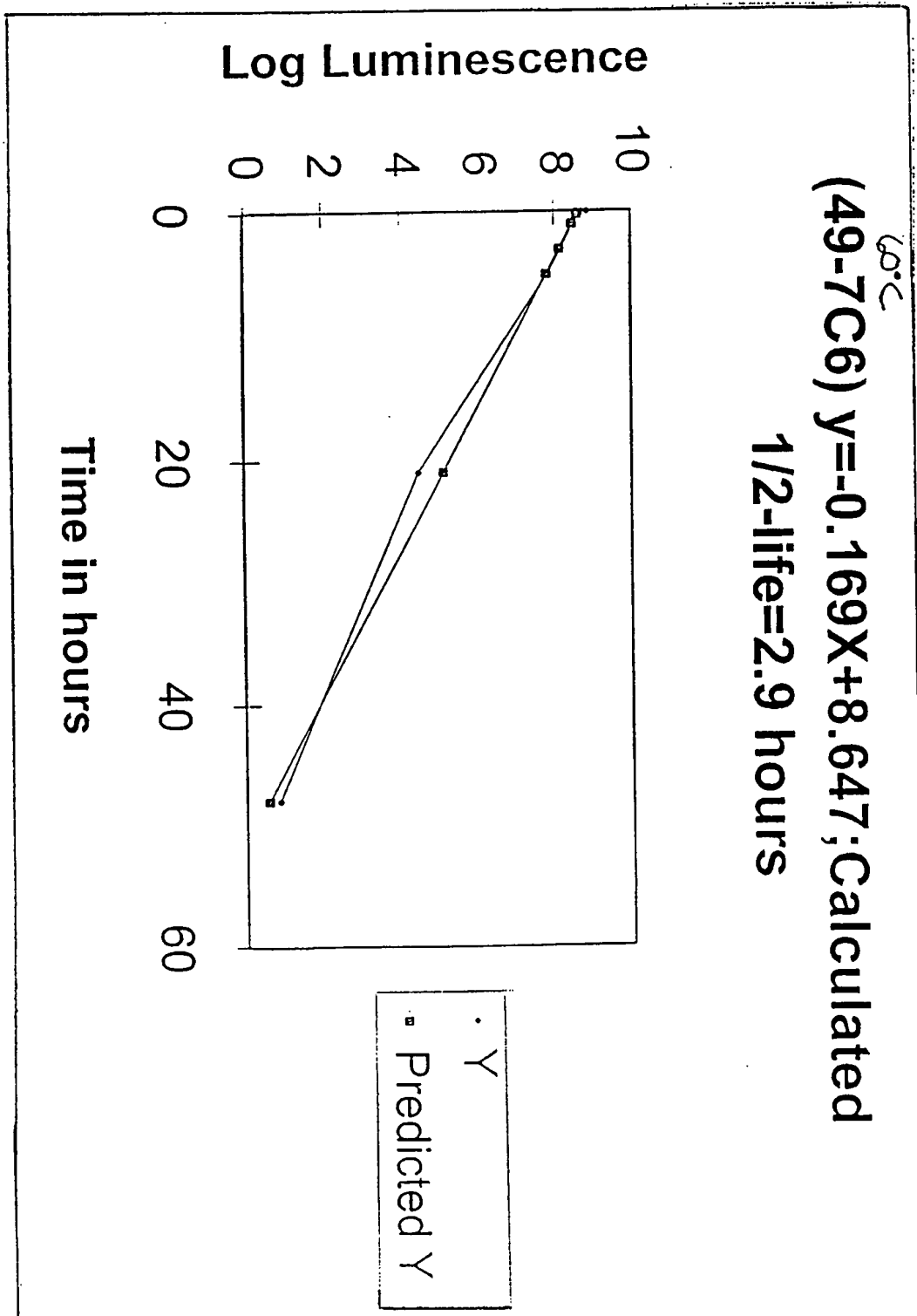




Fig 8



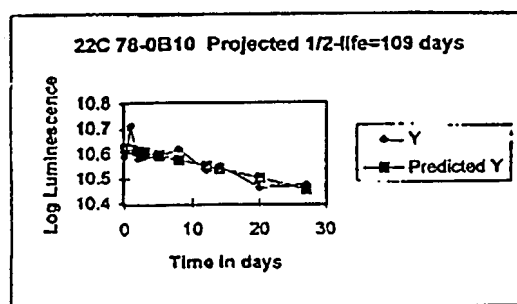


Fig. 9

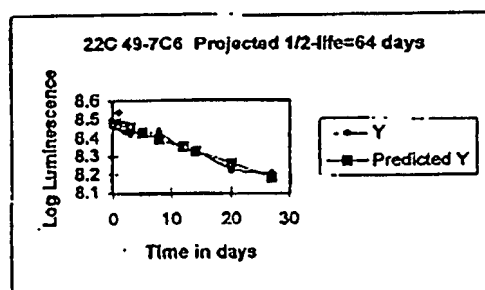
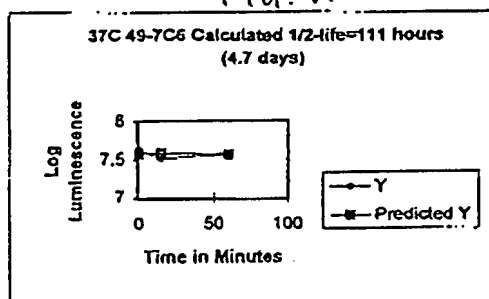


FIG. 10

Fig. 11



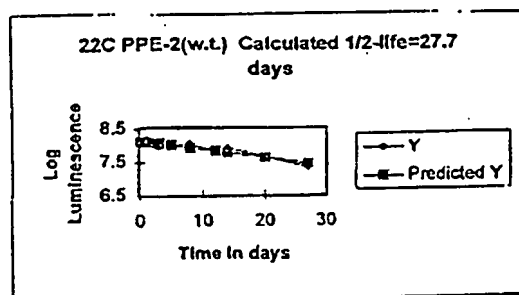


Fig 12

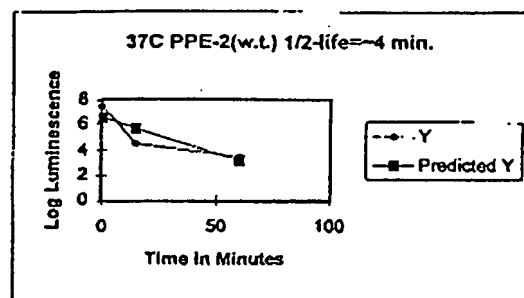
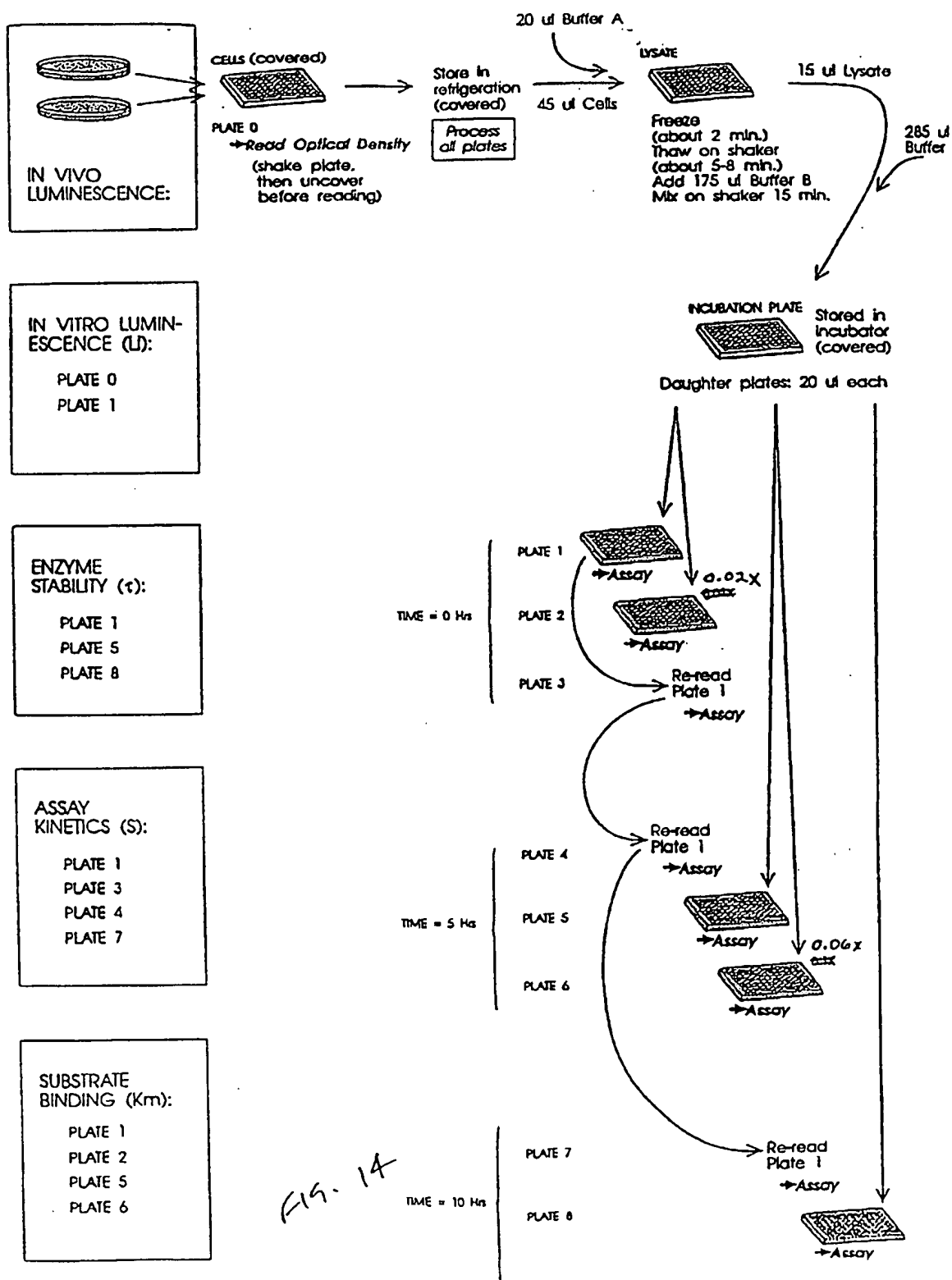


Fig 13



## PROMEGA - TABLE TOP LAYOUT

T265 ROBOT ON 3M TRACK

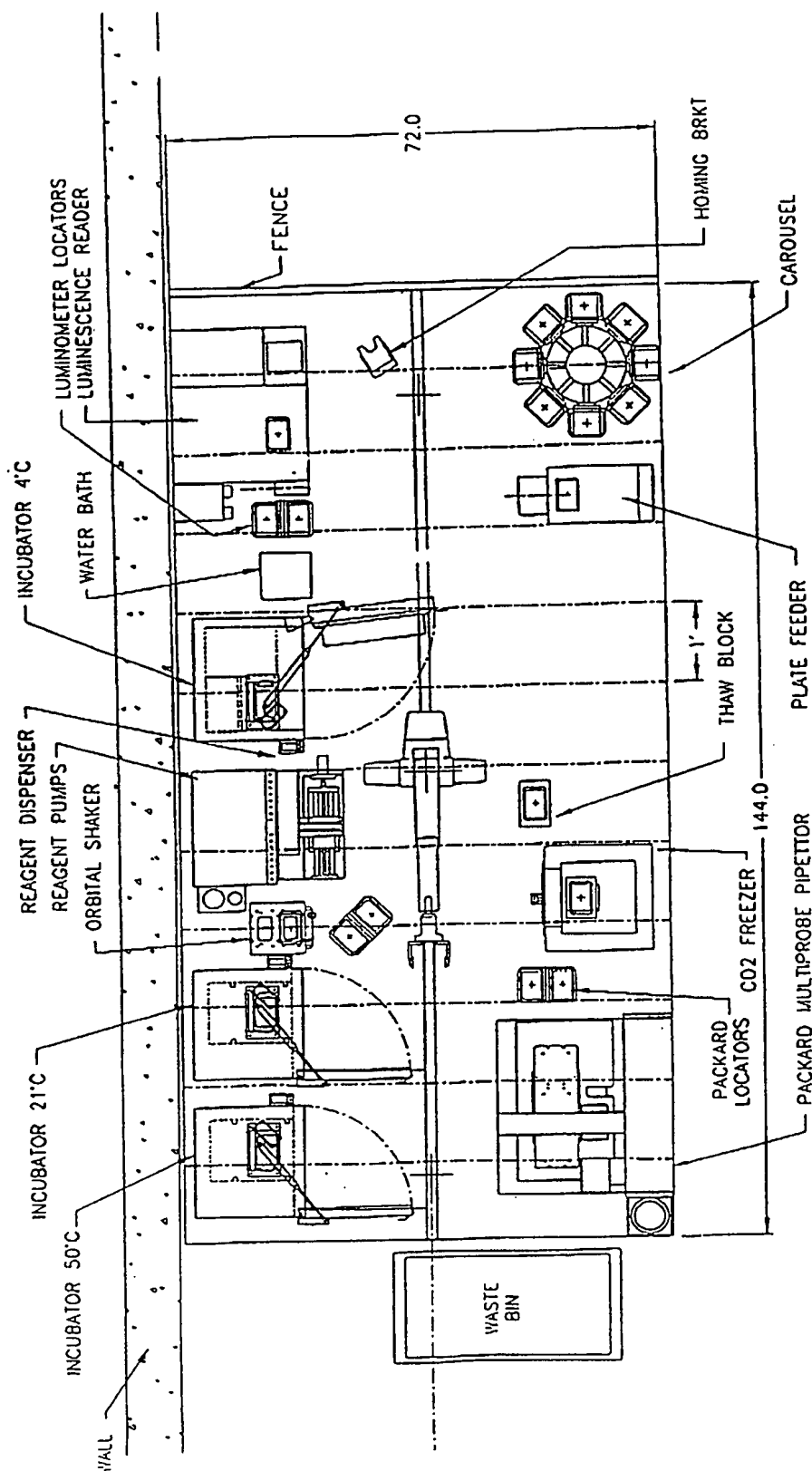
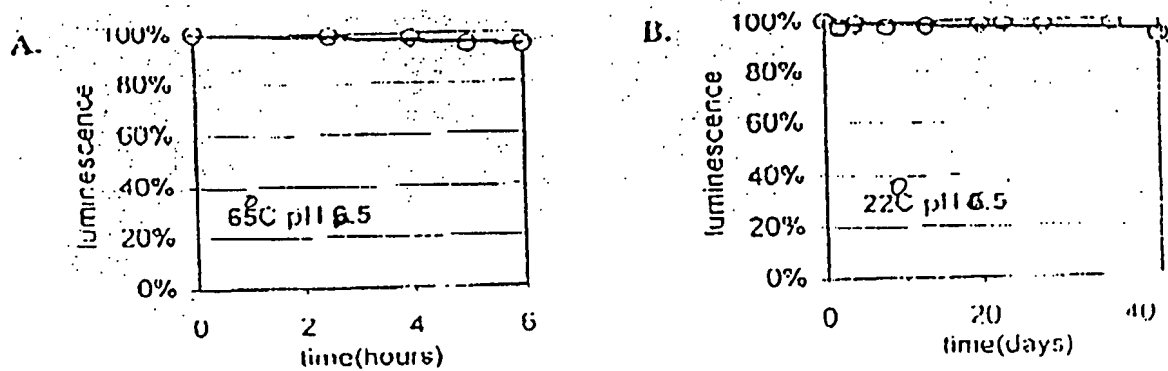


FIG. 15



FIGURE 16



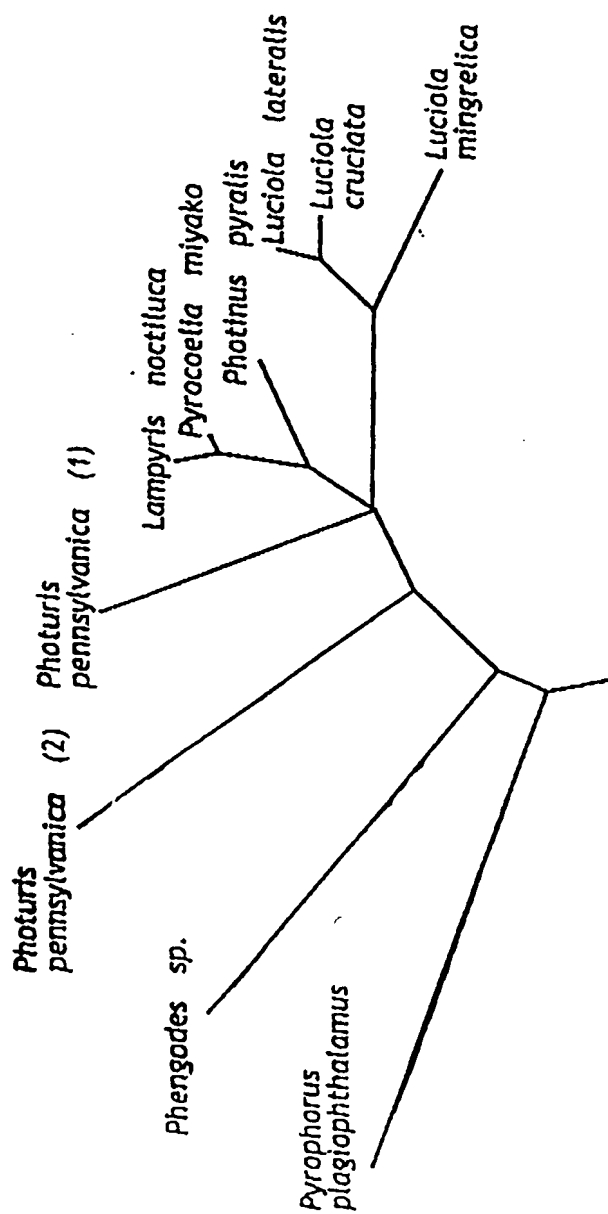


Figure 17

FIGURE 18A



FIGURE 18B

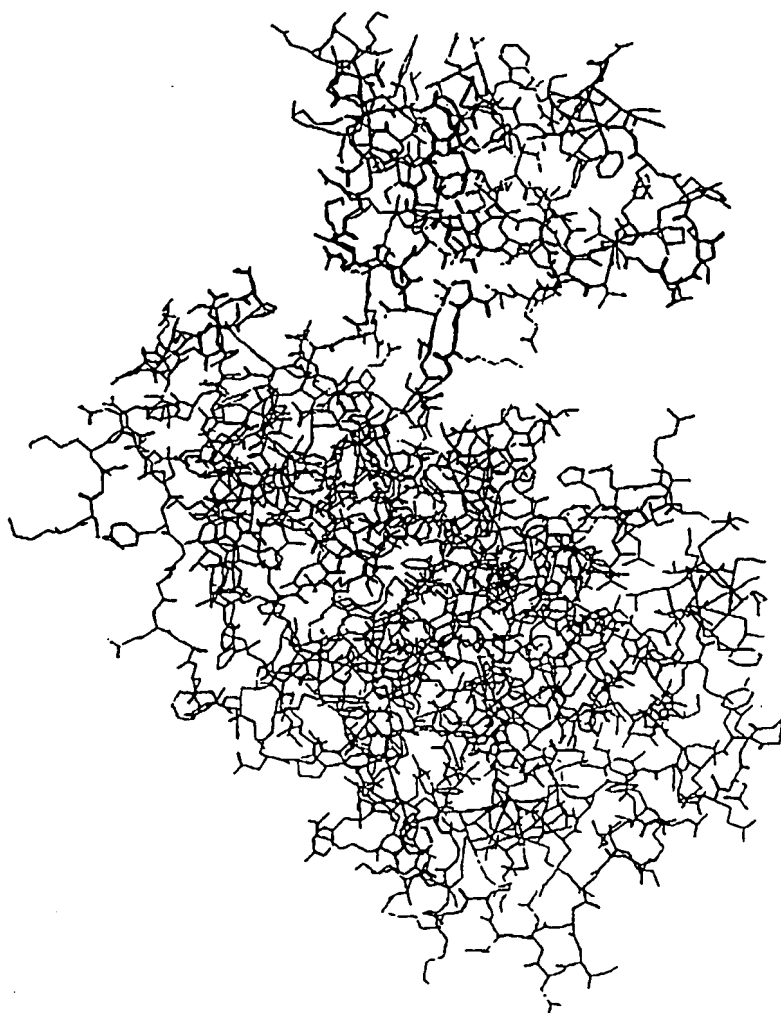


FIGURE 18C

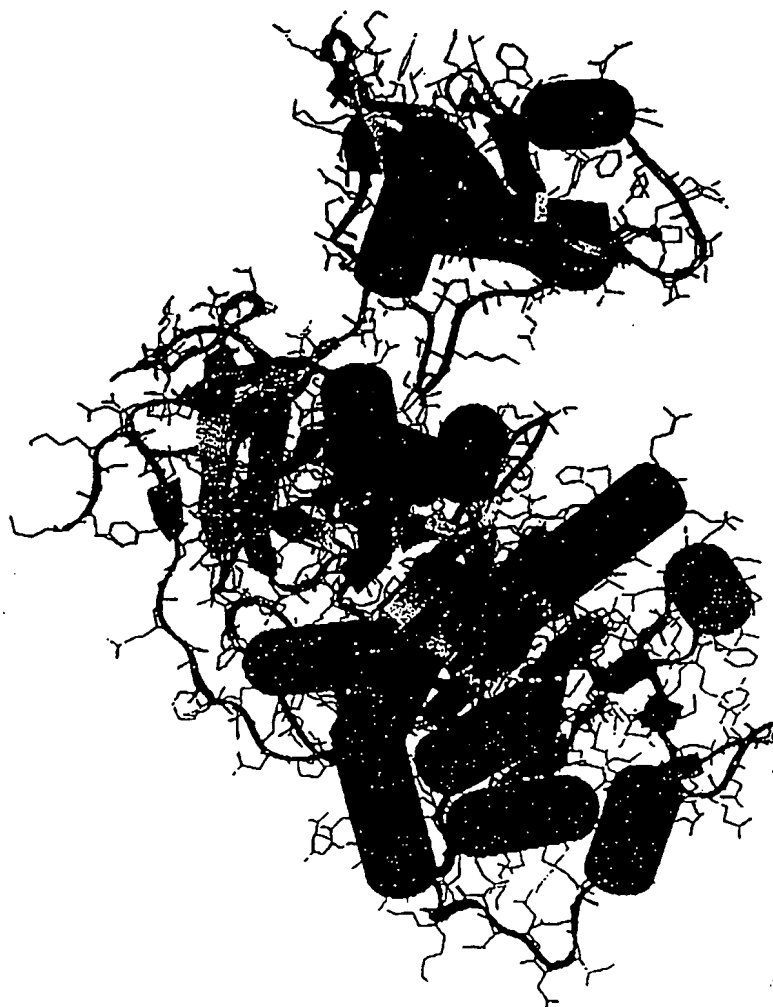


FIGURE 19

1					50
Lcr	MENMENDE.N	IVVGPKPFYP	IEEGSAGTQL	RKYMERYAKL	.GAIAFTNAV
Lla	MENMENDE.N	IVYGPEPFYP	IEEGSAGAQL	RKYMORYAKL	.GAIAFTNAL
Lmi	ME.MEKEE.N	VVYGPLPFYP	IEEGSAGIQL	HKYMHQYAKL	.GAIAFSNAL
Pmi	...MEDDSKH	IMHGHRSIL	WEDGTAGEQL	HKAMKRYAQV	PGTIAFTDAH
Ppy	...MED.AKN	IKKGPAFFYP	LEDGTAGEQL	HKAMKRYALV	PGTIAFTDAH
Lno	...MED.AKN	IMHGPAFFYP	LEDGTAGEQL	HKAMKRYAQV	PGTIAFTDAH
Ppe1	...MSI.ENN	ILIGPPPYYP	LEEGTAGEQL	HRAISRYAAV	PGTLAYTDVH
Phg	MIKME..EEH	VMPGAMPRL	LFEGTAGQQL	HRALYKHSYF	PE..AIVDSH
GR	...MMKREKN	VVYGPEPLHP	LEDLTAGEML	FRALRKHSHL	PQ..ALVDVY
YG	...MMKREKN	VVYGPEPLHP	LEDLTAGEML	FRALRKHSHL	PQ..ALVDVF
Ppe2	...MED..KN	ILYGPEPFYP	LADGTAGEQM	FYALSRYADI	SGCIALTNAH
49-7c6					
78-0b10	A			D	P
90-1b5	A		E	L D	P
Cons	---M-----	---G-----	-----AG---	-----A-----	
51					100
Lcr	TGVDYSYAEY	LEKSCCLGKA	LQNYGLVVDG	RIALCSENCE	EFFIPVIAGL
Lla	TGVDYTYAEY	LEKSCCLGEA	LKNYGLVVDG	RIALCSENCE	EFFIPVLAGL
Lmi	TGVDISYQEY	FDITCRLAEA	MKNFGMKPEE	HIALCSENCE	EFFIPVLAGL
Pmi	AEVNITYSEY	FEMSCRLAET	MKRYGLGLQH	HIAVCSETSL	QFFMPVCGAL
Ppy	IEVNITYAEY	FEMSVRLAEA	MKRYGLNTNH	RIVVCSENSL	QFFMPVLAGL
Lno	AEVNITYSEY	FEMACRLAET	MKRYGLGLQH	HIAVCSENSL	QFFMPVCGAL
Ppe1	TELEVITYKEF	LDVTCRLAEA	MKNYGLGLQH	TISVCSENCV	QFFMPICAAAL
Phg	THEIISYAKI	LDMSCLRAVS	FQKYGLTQNN	IIGICSENNL	NFFNPVIAAF
GR	GEEWISYKEF	FETTCLLAQS	LHNCGYKMSD	VVSICAENNK	RFFVPPIIAAW
YG	GDESLSYKEF	FEATCLLAQS	LHNCGYKMND	VVSICAENNK	RFFVPPIIAAW
Ppe2	TKENVLYEEF	LKLSCRLAES	FKKYGLKQND	TIAVCSENGL	QFFLPLIASL
49-7C6					I
78-0b10					V
90-1b5					V
Cons	-----Y---	-----L---	-----G-----	-----C-E---	-FF-P-----
101					150
Lcr	FIGVGVAPT	EIYTLRELH	SLGISKPTIV	FSSKKGLDKV	ITVQKTVTTI
Lla	FIGVGVAPT	EIYTLRELH	SLGISKPTIV	FSSKKGLDKV	ITVQKTVATI
Lmi	YIGVAVAPT	EIYTLRELH	SLGIAQPTIV	FSSRKGLPKV	LEVQKTVTCI
Pmi	FIGVGVAPT	DIYNERELYN	SLFISQPTIV	FCSKRALQKI	LGVQKKLPVI
Ppy	FIGVAVAPT	DIYNERELYN	SMNISQPTVV	FVSKKGLQKI	LNQVKKLPPI
Lno	FIGVGVAPT	DIYNERELYN	SLSISQPTIV	SCSKRALQKI	LGVQKKLPPI
Ppe1	YVG VATAPT	DIYNERELYN	SLSISQPTVV	FTSRNSLQKI	LGVQSRLPPI
Phg	YLGITVATVN	DTYTDRELSE	TLNITKPQML	FCSKQSLPIV	MKTMKIMPYV
GR	YIGMIVAPVN	EGYIPDELCK	VMGISRPQLV	FCTKNILNKV	LEVQSRTDPI
YG	YIGMIVAPVN	ESYIPDELCK	VMGISRPQIV	FCTKNILNKV	LEVQSRTNFI
Ppe2	YLGIIAAPVS	DKYIERELIH	SLGIVKPRII	FCSKNTFQKV	LNQVSKLKYV
49-7C6					
78-0b10					S
90-1b5	V	N		V	Q SI
Cons	--G---A---	--Y---EL--	---I---P---	-----	-----

22 / 59

151

200

Lcr KTIVILDSKV DYRGYQCLDT FIKRNTPPGF QASSFKTVEV .DRKEQVALI  
 Lla KTIVILDSKV DYRGYQSDMN FIKKNTPPGF KGSSFKTVEV .NRKEQVALI  
 Lmi KKIVILDSKV NFGGHDCMET FIKKHVELGF QPSSEVPIDV KNRKQHVALL  
 Pmi QKIVILDSRE DYMGKQSMYS FIESHLPAGF NEYDYIPDSF .DRETATALI  
 Ppy QKIIIMDSKT DYQGFQSMYT FVTSHLPPGF NEYDFVPESF .DRDKTIALI  
 Lno QKIVILDSRE DYMGKQSMYS FIESHLPAGF NEYDYIPDSF .DRETATALI  
 Ppe1 KKIIILDGKK DYLGYSMQS FMKEHVPANF NVSAFKPLSF .DLDR.VACI  
 Phg QKLLIIDSMT DIGGIECVHS FVSRYTDEHF DPLKFVPLDF .DPREQVALI  
 GR KRIIILDAVE NIHGCESLPN FISRYSDGNI A..NFKPLHY .DPVEQVAAI  
 YG KRIIILDVE NIHGCESLPN FISRYSDGNI A..NFKPLHY .DPVEQVAAI  
 Ppe2 ETIIILDLE DLGGYQCLNN FISQNSDINL DVKKFKPNSF .NRDDQVALV

49-7C6 Y  
 78-0b10 Y  
 90-1b5 Y L(I)

Cons ----I-D--- ---G----- F-----A--  
 201 250

Lcr MNSSGSTGLP KGVQLTHENT VTRFSHARDP IYGNQVSPGT AVLTVPFHH  
 Lla MNSSGSTGLP KGVQLTHENA VTRFSHARDP IYGNQVSPGT AILTVPFHH  
 Lmi MNSSGSTGLP KGVRIHGA VTRFSHARDP IYGNQVSPGT AILTVPFHH  
 Pmi MNSSGSTGLP KGVDLTHMN CVRFSHARDP IYGNQVSPGT AILTVPFHH  
 Ppy MNSSGSTGLP KGVDLTHMN CVRFSHARDP IYGNQVSPGT AILTVPFHH  
 Lno MNSSGSTGLP KGVDLTHMN CVRFSHARDP IYGNQVSPGT AILTVPFHH  
 Ppe1 MNSSGSTGLP KGVPIHHRNT IYRFSHARDP IYGNQVSPGT AILTVPFHH  
 Phg MTSSGTTGLP KGVMLTHRNI CVRFVHSRDP LFGTRFIPET SILSLVPFHH  
 GR LCSSGTTGLP KGVMLTHRNI CVRFVHSRDP LFGTRFIPET SILSLVPFHH  
 YG LCSSGTTGLP KGVMLTHRNI CVRFVHSRDP LFGTRFIPET SILSLVPFHH  
 Ppe2 MFSSGTTGVS KGVMLTHRNI CVRFVHSRDP LFGTRFIPET SILSLVPFHH

49-7c6 LA  
 78-0b10 LA  
 90-1b5 LP LA

Cons --SSG-TG-- KGV---H--- --R--H--DP --G---P-- --L---PF-H  
 251 300

Lcr GFGMFTTLGY LICGFRVVM LTKFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Lla GFGMFTTLGY LTCGFRIVML TKFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Lmi GFGMFTTLGY FACGYRVVM LTKFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Pmi VFQMFTTLGY LTCGFRIVML YRFEEELFLR SLQDYKIQSA LLVPTLFSFF  
 Ppy GFGMFTTLGY LICGFRVVM LTKFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Lno GFGMFTTLGY LTCGFRIVML YRFEEELFLR SLQDYKIQSA LLVPTLFSFF  
 Ppe1 AFGTFTNLGY LICGFRVVM LTKFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Phg AFGMFTTLGY FIVGLKIVMM KRFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 GR AFGFSINLGY FMVGLRVIML RRFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 YG AFGFSINLGY FMVGLRVIML RRFDEETFLK TLQDYKCTSV ILVPTLFAIL  
 Ppe2 GFGMFTTLGY LTCGFRIVML YRFEEELFLR SLQDYKIQSA LLVPTLFAIL

49-7c6 M V L  
 78-0b10 M V L  
 90-1b5 M V R E L

Cons -F-----I-Y ---G----- -F-----FL- --Q-Y----- -P-----

301 350

Lcr NKSELLNKYD LSNLVEIASG GAPLSKEVGE AVARRFNLPG VRQGYGLTET  
 Lla NRSELLDKYD LSNLVEIASG GAPLSKEIGE AVARRFNLPG VRQGYGLTET  
 Lmi NKSELIDKFD LSNLVEIASG GAPLAKEVGE AVARRFNLPG VRQGYGLTET  
 Pmi AKSTLVDKYD LSNLHEIASG GAPLAKEVGE AVAKRFKLPG IRQGYGLTET  
 Ppy AKSTLIDKYD LSNLHEIASG GAPLSKEVGE AVAKRFHLPG IRQGYGLTET  
 Lno AKSTLVDKYD LSNLHEIASG GAPLAKEVGE AVAKRFKLPG IRQGYGLTET  
 Ppe1 AKNPLVDKYD LSNLHEIASG GAPLSKEISE IAAKRFKLPG IRQGYGLTET  
 Phg AKSHLVDKYD LSSIKEIATG GAPLGPALAN AVAKRLKLG IIGYGLTET  
 GR SKSPLVDKYD LSSLRELCCG AAPLAKEVAE IAVKRLNLPG IRCGFGLTES  
 YG SKSPLVDKYD LSSLRELCCG AAPLAKEVAE VAVKRLNLPG IRCGFGLTES  
 Ppe2 AKSALVEKYD LSHLKEIASG GAPLSKEIGE MVKKRFKLPF VRQGYGLTET

49-7C6  
 78-Ob10  
 90-1b5

Cons ----L--K-D LS---E---G -APL----- ----R--L-- ---G-GLTE-

351 400

Lcr TSAIIITPEG DDKPGASGKV VPLFKAKVID LDTKKSLGPN RRGEVCVKGP  
 Lla TSAIIITPEG DDKPGASGKV VPLFKAKVID LDTKKTLGPN RRGEVCVKGP  
 Lmi TSAFIITPEG DDKPGASGKV VPLFKVKVID LDTKKTLGPN RRGEICVKGP  
 Pmi TSAIIITPEG DDKPGACGKV VPFFTAIVD LDTGKTLGPN QRGELCVKGP  
 Ppy TSAIIITPEG DDKPGAVGKV VPFFFAKVID LDTGKTLGPN QRGELCVKGP  
 Lno TSAIIITPEG DDKPGACGKV VPFFSAKVID LDTGKTLGPN QRGELCVKGP  
 Ppe1 TCAIVITAEG EFKLGAAGKV VPFFSLKVID LNTGKKNLGN ERGEICFKGP  
 Phg CCAVLITPHN KIKTGSTGV LPYVTAKIVD TKTGKNLGN QVGELCFKSD  
 GR TSANIHSRDL EFKSGSLGRV TPLMAAKIAD RETGKALGN QVGELCIKGP  
 YG TSANIHSRDL EFKSGSLGRV TPLMAAKIAD RETGKALGN QVGELCVKGP  
 Ppe2 TSAVLITPDT DVPRGSTGKI VPFHAVKVID PTTGKILGN ETGELYFKGP

49-7C6 NN P A  
 78-Ob10 xx A P P  
 90-1b5 xx A

Cons --A----- ---G--G-- -P----K--D --T-K-LG-N --GE-----

401 450

Lcr MLMKGYVNNP EATKELIDEE GWLHTGDIGY YDEEKHFFIV DRLKSLIKYK  
 Lla MLMKGYVDNP EATREIIDE E GWLHTGDIGY YDEEKHFFIV DRLKSLIKYK  
 Lmi SLMLGYSNNP EATRETIDEE GWLHTGDIGY YDEDEHFFIV DRLKSLIKYK  
 Pmi MIMKGYVNNP EATNALIDKO GWLHSGDIAY YDKDGHFFIV DRLKSLIKYK  
 Ppy MIMSGYVNNP EATNALIDKO GWLHSGDIAY YDKDGHFFIV DRLKSLIKYK  
 Lno MIMKGYVNNP EATSALIDKO GWLHSGDIAY YDKDGHFFIV DRLKSLIKYK  
 Ppe1 MIMKGYINNP EATRELIDEE GWIHSGDIGY FDEDGHVYIV DRLKSLIKYK  
 Phg IIMKGYQNE EETRLVIDKO GWLHSGDIGY YDTGDNFHV DRLKELIKYK  
 GR MVSKGYVNNV EATKEAIDDD GWLHSGDFGY YDEDEHFYVV DRYKELIKYK  
 YG MVSKGYVNNV EATKEAIDDD GWLHSGDFGY YDEDEHFYVV DRYKELIKYK  
 Ppe2 MIMKSYNNE EATKAIINKO GWLRSGLIAY YDNDGHFYIV DRLKSLIKYK

49-7C6 G DN  
 78-Ob10 G DN  
 90-1b5 G

Cons -----Y--N- E-T---i--- GW---GD--Y -D-----V DR-K-LIKYK



451 500

Lcr GYQVPPAELE SVLLQHPSIF DAGVAGVPDP VAGELPGAVV VLESGKNMTE  
 Lla GYQVPPAELE SVLLQHPSIF DAGVAGVPDP IAGELPGAVV VLEKGKSMTE  
 Lmi GYQVPPAELE SVLLQHPSIF DAGVAGVPDP DAGELPGAVV VMEKGKMTTE  
 Pmi GYQVPPAELE SILLQHPSIF DAGVAGIPDP DAGELPAAVV VLEEGKMTTE  
 Ppy GYQVAPAELE SILLQHPSIF DAGVAGLPDD DAGELPAAVV VLEHGKMTTE  
 Lno GYQVPPAELE SILLQHPSIF DAGVAGIPDP DAGELPAAVV VLEEGKMTTE  
 Ppel GYQVPPAELE ALLLQHPSIF DAGVAGVPDE VAGDLPGAVV VLKEGKSITE  
 Phg AYQVAPAELE ALLLQHPYIA DAGVTGIPDE EAGELPAACV VLEPGKMTTE  
 GR GSQVAPAELE EILLKNPCIR DVAVVGIPDL EAGELPSAFV VIQPGKEITA  
 YG GSQVAPAELE EILLKNPCIR DVAVVGIPDL EAGELPSAFV VKQPGKEITA  
 Ppe2 GYQVAPAEIE GILLQHPYIV DAGVTGIPDE AAGELPAAGV VVQTGKYLNE

49-7c6  
 78-0b10  
 90-1b5

Cons --QV-PAE-E --LL--P-I- D--V-G-PD- -AG-LP-A-V V---GK----

501 550

Lcr KEVMDYVASQ VSNAKRLRGG VRFVDEVPGK LTGKIDGRA. IREILKKPV.  
 Lla KEVMDYVASQ VSNAKRLRGG VRFVDEVPGK LTGKIDGKA. IREILKKPV.  
 Lmi KEIVDYVNSQ VVNHKRLRGG VRFVDEVPGK LTGKIDAKV. IREILKKPQ.  
 Pmi QEVMDYVAGQ VTASKRLRGG VKFVDEVPGK LTGKIDSRK. IREILTMGQK  
 Ppy KEIVDYVASQ VTTAKKL RGG VV FVDEVPGK LTGKLDARK. IREILIKAKK  
 Lno QEVMDYVAGQ VTASKRLRGG VKFVDEVPGK LTGKIDGRK. IREILMMGKK  
 Ppel KEIQDYVAGQ VTSSKKLRGG VEFVKEVPKG FTGKIDTRK. IKEILIKAQK  
 Phg KEVMDYIAER VTPTKRLRGG VLFVNNIPKG ATGKLV RTE. LRRLLTQRA.  
 GR KEVYDYLAER VSHTKYL RGG VRFVDSIPRN VTGKITRKEL LKQLEKS..  
 YG KEVYDYLAER VSHTKYL RGG VRFVDSIPRN VTGKITRKEL LKQLEKS..  
 Ppe2 QIVQNFVSSQ VSTAKWLRGG VKFLDEIPKG STGKIDRKV. LRQMFEEKH..

49-7c6  
 78-0b10 D  
 90-1b5 DY A

Cons ----- V---K-LRGG V-F----P-- -TGK-----

551

Lcr .....AKM  
 Lla .....AKM  
 Lmi .....AKM  
 Pmi .....SKL  
 Ppy G...GKSKL  
 Lno .....SKL  
 Ppel GKSKSKAKL  
 Phg .....AKL  
 GR .....SKL  
 YG .....SKL  
 Ppe2 .....KSKL

49-7c6 TNG  
 78-0b10 TNG  
 90-1b5 TNG

## Key:

Lcr: *Luciola cruciata*  
Lla: *Luciola lateralis*  
Lmi: *Luciola mingrelica*  
Pmi: *Pyrocoelia miyako*  
Ppy: *Photinus pyralis*  
Lno: *Lampyris noctiluca*  
Ppe-1: *Photuris pennsylvanica* (1)  
Phg: *Phengodes* sp.  
Gr: *Pyrophorus plagiophthalmus* (green)  
YG: *Pyrophorus plagiophthalmus* (yellow green)  
Ppe-2: *Photuris pennsylvanica* (2)

FIGURE 20

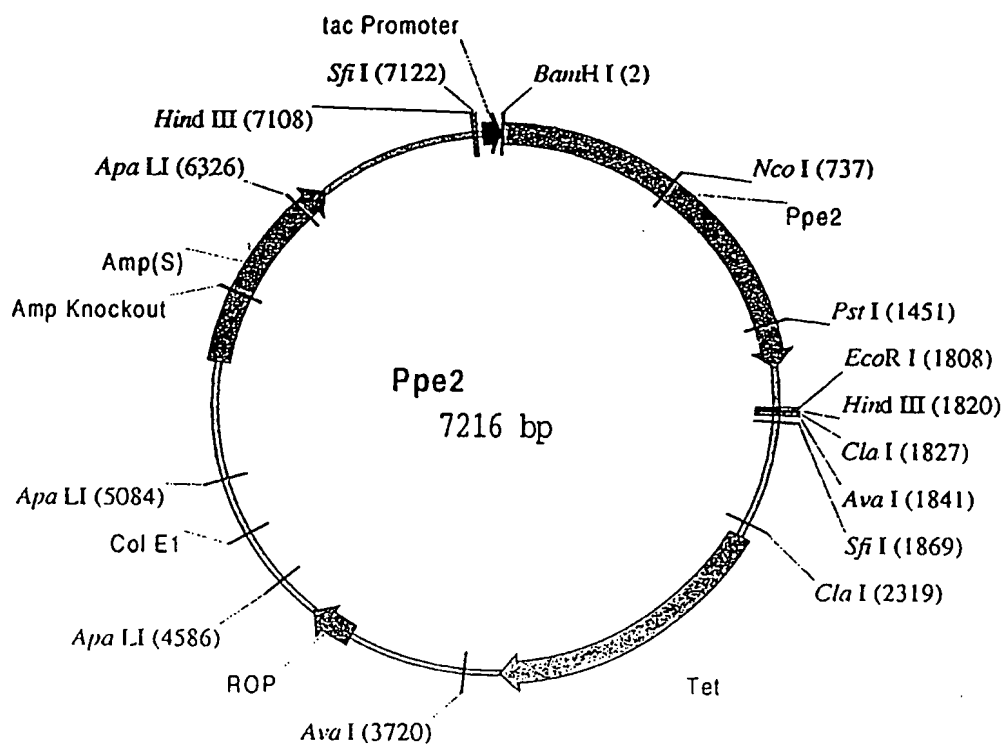
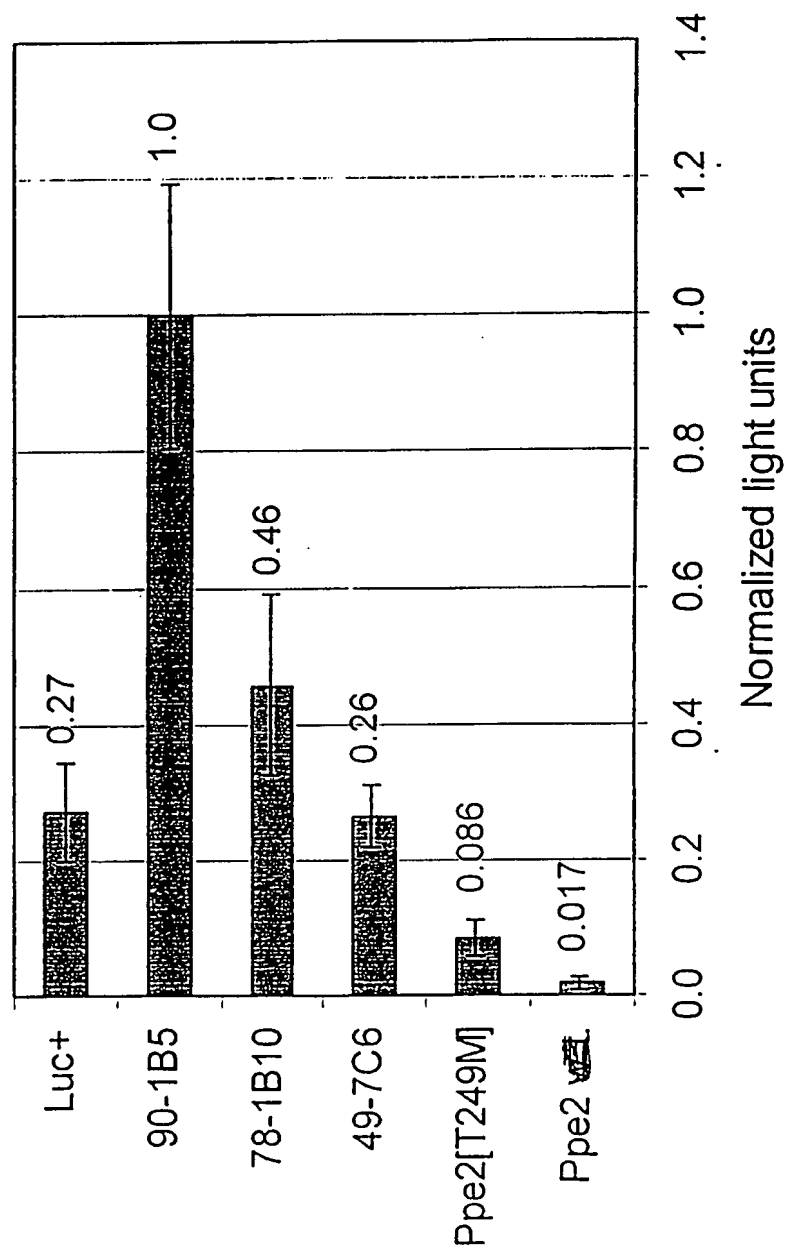


FIGURE 21

## Luminescence

Expression from colonies of *E. coli*



## FIGURE 22

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTCCCTTCTATAATTGCATCATTGTATCTTGG  
AATAATTGCAGCAGCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAACCACGCATAATTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATCTTTTAATCGAGACGATCAGGTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTAAC  
GGTAATACCTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCGAGTGTCTTAATGCACACGTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGA AAAAACGGTTAAATTAATCTTGTGAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACAATGACGTCAGACC  
GGGATCAACTGGTAAAAATAGTACCATTTACGCTGTAAAGTTGTCGATCTACAACAGG  
AAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
AGGTTATTATAATAATGAAGAAGCTACTAAGCAATTATTAACAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAATCGTACA  
AAATTTGTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTT  
GGATGAAATCCCAAAGGATCAACTGGAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACACCAATGGG\*

## FIGURE 23

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTGTTAAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAACAAACGACACAAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTCTCTCTATAATTCATCATTTGATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCACGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAACTAAATTAATAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACTTTATTCTCAAAATTCGATATTAATCTGGACGTAAA  
AAAATTTAAACCATATTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCATGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAGTACTTTACTTGTACCAACATTAATGGCATTTTTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATGTCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTAAATTAACTTTGTCAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACCAATGACGTCAGACC  
GGGATCACTGGTAAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCCTACACAGG  
AAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAATAATCTAAACGAACAAATCGTACA  
AAATTTGTTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATCCCAAAGGATCAACTGGAATAATGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACACCAATGGG\*

## FIGURE 24

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTCAAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTCTTCTCTATAAATGCATCATTTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCCAGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAAATATGTAGAAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACCTTATTCTCAAAATCCGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGTACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGAAAAACGGTTTAAATTAACTTTGTCAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACAATGACGTGAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
AGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTACCAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAAATTAATATAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAATTTTT  
GGATGAAATCCCAAGGATCAACTGAAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACCAATGGG\*

FIGURE 25

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTTACGCATTATCTCGTTATGCAGATATTTTCAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAAAT  
GTCGTGTCGTTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTAGCGAAAATGGTTTGCAATTTTCTTCCTATAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAACTAAATTAATAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACCTTTATTCTCAAAATTCGGATATTAATCTTGACGTAAA  
AAAATTTAAACCATATTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGATATTGTGCACGATT  
TTCTATTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAAATGTCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTTGGGGAGATGGTGAAAAACCGTTTAAATTAACCTTTGTGAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACCGGAACAATGACGTCAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
AGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTAACAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTTT  
GGATGAAATTTCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACCAATGGG\*



FIGURE 26

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTTCAAGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAATGGTTTGCATTTTCTCTTCTATATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCACGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACTTTATTTCTCAAAATTCGGATATTAACTTTGACGTAAA  
AAAATTTAAACCATATTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTC  
TGGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCATGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGAGGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAAATGGCATTTTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATGCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTTGGGGAGATGGTGAAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAACAATGACGTGAGACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGGCGACATGATAATGAA  
AGGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTAACAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTT  
GGATGAAATTTCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACCAATGGG\*

## FIGURE 27

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIACVSENGLOFFLPPIASLYLGIIAAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSENRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMMHTFEELFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPNDVVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEEATKAIINKDGWLR  
SGDIAYYDNDGHEYIVDRLKSLIKYQYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

## FIGURE 28

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEELLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPITIASLYLGIIAAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSAKDPFNGAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEATKAIINKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

## FIGURE 29

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPITIASLYLGIIAAPVSDKYIERELIHS LG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVVRFS LAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKREKLN FVRQGYGLTETTS AVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEEATKAIITKDGWLR  
SGDIAYYDNDGHFYIVDRLKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

## FIGURE 30

DPMEDKNILYGPEFFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIIVCSENGLQFFLPPIIASLYLGIIAAPVSDKYIERELIHSIG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSIAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKREKLNFRQGYGLTETTSAVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNNEATKAIINKDGWLR  
SGDIAYYDNDGHFYIVDRKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

## FIGURE 31

DPMEDKNILYGPEPFYPLADGTAGEQMFDAISRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPPIIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKYVETIIILDNLNEDLGGYQCLNNFISONSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSHAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFQSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPNNDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKGYNNEEATKAIINKDGWLR  
SGDIAYYDNDGHFYIVDRKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHTNG

FIGURE 32

GGATCCAATGGCAGATAAAAAATATTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTTCAAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCATTTTTCCTTCCGTAATTGCATCATTGTATCTTGGA  
ATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGTT  
ATTGTAAAACCGCATAAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGTA  
AAATCTAAATTAATCTGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGGA  
GGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAAA  
AAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTCT  
GGTACAACCTGGTGTTCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATTT  
TCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCCACGACAGCAATTTAACG  
GTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGGA  
TTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACCTATTCTACAATCATTACAAGAT  
TATAAAGTGGAAAGTACTTTACTTGTACCAACATTAAATGGCATTCTTGCAAAAAGTGCA  
TTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAAATTGCATCTGGTGGCGCACCTTTA  
TCAAAAAGAAATGGGGAGATGGTGAACCAACGTTTAAATTAACCTTTGTGAGCAAGGG  
TATGGATTAACAGAAACCACTTCGGCTGTTTTAATTACACCGAAAXXXXXGCCAGACCG  
GGATCAACTGGTAAATAGTACCATTTTACGCTGTTAAAGTTGTCGATCCTACAACAGGA  
AAAATTTGGGGCCAAATGAACCTGGAGAATTGTATTTAAAGGCGCCATGATAATGAAG  
GGTTATTATAAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCCG  
TCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAAG  
TCATTAATTAAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATCTCTTA  
CAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATTCGGATGAAGCCGCGGGCGAG  
CTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACAA  
GATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTTG  
GATGAAATCCCAAAGGATCAACTGGAATAATTGACAGAAAAGTGTAAAGACAAATGTTT  
GAAAACACACCAATGGG\*

## FIGURE 33

GGATCCAATGGCAGATAAAAAATATTTTATATGGGCCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTACGCATTATCTCGTTATGCAGATATTTCAGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTAAAAAT  
GTCGTGTCGTTTACGGGAAAGTTTAAAAAGTATGGATTAAAAACAAACGACACAATAGC  
GGTGTGTAGCGAAATGGTTTGCAATTTTCTCTCCTGTAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCCGCATAATTTTGTCTCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAACTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAACTTTGACGTAAA  
AAAATTTAAACCATATTCTTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
TGGTACAACCTGGTGTCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTCAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAAXXXXXGTACAGAC  
GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTGCGATCTACAAACAG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTAAAGGCGACATGATAATGAA  
AGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAAATATCTAAACGAACAAATCGTACA  
AAATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGGGGTGGGGTGAAATTTT  
GGATGAAATCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACCCAATGGG\*



## FIGURE 34

GGATCCAATGGCAGATAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTCCTGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTAAATTT  
GTCGTGTCGTTAGCGGAAAGTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAATGGTTGCAATATTCCTTCCTGTAATTGCATCATTTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAATTTTGTCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACTTTATTTCTCAAAATTCGATATTAATCTTGACGTAAA  
AAAATTTAAACCAAATCTTTTAAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
TGGTACAACCTGGTGTCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTATTGCAAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTATCGCACTTAAAAGAAATGCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTTGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTGAGGCAAGG  
GTATGGATTAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAAxxxxxGCCAGACC  
GGGATCAACTGGTAAAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGGCGCCATGATAATGAA  
GGGTATTATAAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AAATTTGTTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATTCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTAAAGACAATGTT  
TGAAAAACACCAATGGG\*

## FIGURE 35

GGATCCAATGGCAGATAAAAAATATTTATATGGGCCCGAACCATTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATCCCGGATG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTTGCAATTTTCTTCTCTGTAATTGCATCATTGTATCTTGG  
AATAATTGCAGCACCTGTTAGTGATAAATACGTTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAATTTTTTGCTCCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATATGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACTTTATTTCTCAAAATCCGATAGTAATCTGGACGTAAA  
AAAATTTAAACCAAATCTTTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTCTTC  
TGGTACAACCTGGTGTCCGAAGGGAGTCATGCTAACTCACAAGAATTATTGTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATTTTAAAC  
GGTAATACCTTTCCACCATGGTTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAAATTGCATCTGGTGGCGCACCTTT  
ATCAAAAAGAAATTGGGGAGATGGTGAAAAAACGGTTTAAATTAAGTCTTGTGAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTTAATTACACCGAAAxxxxxxGCCAGACC  
GGGATCAACTGGTAAATAGTACCATTTCACGCTGTTAAAGTTGTCGATCCTACAACAGG  
AAAAATTTGGGGCCAAATGAACCTGGAGAATTGTATTTTAAAGGCGCCATGATAATGAA  
GGGTATTATAAATAATGAAGAAGCTACTAAAGCAATTATTGATAAAGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTTATATTGTGGACAGGCTGAA  
GTCATTAATTAAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAAATCTAAACGAACAAATCGTACA  
AAATTTGTTTCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGGTGAAATTTT  
GGATGAAATTCCTCAAGGATCAACTGGAAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACACCAATGGG+

## FIGURE 36

DPMADKNILYGPEPFYPLADGTAGEQMFDAISRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIIVCSENGLOFFLPVIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKSVEIIILDNLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYETCGFRVVLMTFEEKLFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPK\*\*ARPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGAMIMKGYNNNEATKAIIDNDGWLRS  
GDIAYYDNDGHFYIVDRKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQDFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNGS

## FIGURE 37

DPMADKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLQFFLPVIASLYLGI AAPVSDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKYVETIIILDLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPYSENRDDQVALVMFSSGTTGVPKGVM LTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVL MHTFEEKLFQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTS AVLITPK~~xx~~VRPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGDMIMKGYNN EEA TKAIIDKGWLRS  
GDIAYYDNDGHFYIVDR LKSLIKYKGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

## FIGURE 38

DPHADKNILYGPEPFYPLADGTAGEQMFDAISRYADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOQYFLPVIASLYLGIIAAPVSDKYIERELIHSIG  
IVKPRIIFCSKNTFQKVLNVKSKLKYVETIIILDLNEDLGGYQCLNNFISONSDINLDVK  
KFKPNSFNRDDQVALVMFSSGTTGVPKGVMLTHKNIVARFSIAKOPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVMHTFEELFLQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPK~~AR~~PG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGAMIMKGYNNNEATKAIIDKDGWLR  
GDIAYYDNDGHFYIVDRLKS LIKYGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

## FIGURE 39

DPMADKNILYGPEPFYPLADGTAGEQMFDAISRYADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOFFLPVIASLYLGI AAPVSDKYVERELIHS LG  
IVKPRIIFCSKNTFQKVLNVKSKLYVETIIILDNLNEDLGGYQCLNNFISONSDSNLDVK  
KFKPNSFNRDDQVALVMFSSGTTGVPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYETCGFRVVLMTFEEKLFQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTS AVLITPK\*\*ARPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGAMIMKGYNNNEATKAIIDKDWLRS  
GDIAYYDNDGHFYIVORLKS LIKYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

FIGURE 40

GGATCCAATGGCAGATAAAAAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGCTGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATCCGGGCTG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTTTAAAAAT  
GTCGTGTCGTTTAGCGGAAAGTTTTAAAAAGTATGGATTAAAAACAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTTGCAATTTTCCCTTCTGTAAATGCATCATTTGTATCTTGG  
AATAATTGTGGCACCTGTTAACGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACCCACGCATAGTTTTTGTCTCCAAGAATACTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAATCTGTAGAACTATTATTATATAGACTTAAATGAAGACTTAGG  
AGGTATCAATGCCTCAACAACCTTTATTTCTCAAAATCCGATATTAACTTGCAGTAAA  
AAAATTTAAACCATATTCTTTAATCGAGACGATCAGGTTGCGTTGATTATGTTTCTTC  
TGGTACAACCTGGTCTGCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTGCACGATT  
TTCTCTTGCAAAAGATCCTACTTTTGGTAACGCAATTAATCCACGACAGCAATTTAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAATGGCATTCTTGCAAAAAGTGC  
ATTAGTTGAAAAGTACGATTTATCGCACTTAAAGAAATGTCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTGGGGAGATGGTGAAAAACGGTTTAAATTAAACTTTGTGAGGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAAXXXXXGCCAGACC  
GGGATCAACTGGTAAATAGTACCATTTCAGCTGTTAAAGTTGTCGATCCTACAAACAGG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTAAAGGCCGATGATAATGAA  
GGGTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCATTAAATTAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATTCCGGATGAAGCCGCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAAATCTAAACGAACAAATCGTACA  
AGATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTT  
GGATGAAATTCCCAAAGGATCAACTGGAAAAATTGACAGAAAAGTGTTAAGACAAATGTT  
TGAAAAACACCAATGGG

## FIGURE 41

DPMADKNILYGPEPFYPLADGTAGEQMFDAISRYADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOFFLPVIASLYLGIIVAPVNDKYIERELIHS LG  
IVKPRIVFCSKNTFQKVLNVKSKLKS VETIIILDNLNEDLGGYQCLNNFISONSDINLDVK  
KFKPYSFNRDDQVALIMFSSGTTGLPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLG YFTCGFRVVLMTFEEKLFLOSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTS AVLITPK~~xxx~~ARPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGPMIMKGYNNNEATKAIIDNDGWLRS  
GDIAYYDNDGHFYIVDRLKSLIKYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQDFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG



FIGURE 42

GGATCCAATGGCAGATAAGAATATTTTATATGGGCCCGAACCATTTTATCCCTTGGGAAGA  
TGGGACGGCTGGAGAACAGATGTTTGACGCATTATCTCGTTATGCAGATATTCGGGGCTG  
CATAGCATTGACAAATGCTCATACAAAAGAAAATGTTTATATGAAGAGTTTCTGAAACT  
GTCGTGTCGTTTAGCGGAAAGTTTAAAAAGTATGGATTAAAAACAAAACGACACAATAGC  
GGTGTGTAGCGAAAATGGTCTGCAATTTTTCCTTCTGTAAATGCATCATTGTATCTTGG  
AATAATGTGGCACCTGTTAACGATAAATACATTGAACGTGAATTAATACACAGTCTTGG  
TATTGTAAAACACGCATAATTTTGTCTCAAGAATACTTTTCAAAAAGTACTGAATGT  
AAAATCTAAATTAAATCTGTAGAACTATTATTATATTAGACTTAAATGAAGACTTAGG  
AGGTTATCAATGCCTCAACAACCTTATTCTCAAAATCCGATATTAACTTGTACGTAAA  
AAAATTTAAACCATATTTCTTTAATCGAGACGATCAGGTTGCGTTGTTAATGTTTCTTC  
TGGTACAACCTGGTCTGCCGAAGGGAGTCATGCTAACTCACAAGAATATTGTTGCACGATT  
TTCTCTTGCaaaAGATCCTACTTTTGGTAACGCAATTAATCCACGACAGCAATTTTAAC  
GGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGG  
ATTCCGAGTTGTTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGA  
TTATAAAGTGGAAAGTACTTTACTTGTACCAACATTAAATGGCATTCTTGC AAAAAGTGC  
ATTAGTTGAAAAGTACGATTATCGCACTTAAAAGAAATTCATCTGGTGGCGCACCTTT  
ATCAAAAGAAATTTGGGGAGATGGTGA AAAAACGGTTTAAATTAACTTTGTCAAGCAAGG  
GTATGGATTAAACAGAAACCACTTCGGCTGTTTAAATTACACCGAAAxxxxxGCCAAACC  
GGGATCAACTGGTAAATAGTACCATTTACGCTGTTAAAGTTGTGATCCTACAACAGG  
AAAAATTTTGGGGCCAAATGAACCTGGAGAATTGTATTTAAAGGCCGATGATAATGAA  
GGGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTGATAATGACGGATGGTTGCG  
CTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAA  
GTCACTGATTAAATATAAAGGTTATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTT  
ACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATTCCGGATGAAGCCCGGGCGA  
GCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACA  
AGATTATGTTGCCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTT  
GGATGAAATCCCAAAGGATCAACTGAAAAATTGACAGAAAAGTGTAAAGACAAATGTT  
TGAAAAACACCAATGGG

## FIGURE 43

DPMADKNILYGPEPFYPLEDGTAGEQMFDAISRYADIPGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOFFLPVIASLYLGIIVAPVNDKYIERELIHSLG  
IVKPRIIFCSKNTFQKVLNVKSKLKSVEITIIIDLDNEDLGGYQCLNNFISQNSDINLDVK  
KFKPTSFNRDDQVALLMFSSGTTGLPKGVMLTHKNIVARFSLAKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVVLMTFEEKLFQSLQDYKVESTLLVPTLMAFLAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFLNFVRQGYGLTETTSAVLITPKAKPG  
STGKIVPFHAVKVVDPTTGKILGPNEPGELYFKGPMIMKGYNNNEEATKAIIDNDGWLRS  
GDIAYYDNDGHFYIVDRLKSLIKYGYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGEL  
PAAGVVVQTGKYLNEQIVQDYVASQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMFE  
KHTNG

FIGURE 44

GGATCCAATGGAAGATAAAAAATATTTTATATGGACCTGAACCATTTTATCCCTGGCTGATGGGACGGCTGGAGAACAG  
ATGTTTTACGCATTATCTCGTTATGCAGATATTTAGGATGCATAGCATTGACAAATGCTCATACAAAAGAAAATGTTT  
TATATGAAGAGTTTTTAAAATTGTCTGTCTGTTAGCGGAAAAGTTTTAAAAAGTATGGATTAAAAACAAACGACACAAT  
AGCGGTGTGTAGCGAAAATGGTTTGCAATTTTTCCTTCCTTAATTGCATCATTGTATCTTGGAATAATTGCAGCACCT  
GTTAGTGATAAATACATTGAACGTGAATTAATACACAGTCTTGGTATTGTAAAACACGCATAATTTTTGTTCCAAGA  
ATACTTTTCAAAAAGTACTGAATGTAAAATCTAAATTAATAATGTAGAACTATTATTATATTAGACTTAAATGAAGA  
CTTAGGAGGTTATCAATGCCTCAACAACTTTATTTCTCAAAATCCGATATTAATCTTGACGTAAAAAATTTAAACCA  
AATTCITTTAATCGAGACGATCAGGTTGCGTTGGTAATGTTTTCTTCTGGTACAACCTGGTGTTCGAAGGGAGTCATGC  
TAACCTACAAGAATATTGTTGCACGATTTTCTCATGCAAAGATCCTACTTTGGTAACGCAATTAATCCAACGACAGC  
AATTTTAACGGTAATACCTTTCCACCATGGTTTGGTATGATGACCACATTAGGATACTTTACTTGTGGATTCCGAGTT  
GCTCTAATGCACACGTTTGAAGAAAACTATTTCTACAATCATTACAAGATTATAAAGTGGAAAGTACTTTACTTGTAC  
CAACATTAAATGGCATTTTTTGCAAAAAGTGCATTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTCATCTGG  
TGGCGCACCTTTATCAAAAGAAATTTGGGGAGATGGTGAAAAACGGTTTAAATTAACCTTTGTGAGGCAAGGGTATGGA  
TTAACAGAAACCACTTCGGCTGTTTTAATTACACCGGACACTGACGTGAGACCGGGATCAACTGGTAAATAGTACCAT  
TTCACGCTGTAAAGTTGTCGATCCTACAACAGGAAAAATTTGGGGCCAAATGAACTGGAGAATTGTATTTTAAAGG  
CGACATGATAATGAAAAGTTATTATAATAATGAAGAAGCTACTAAAGCAATTATTAAACAAAGACGGATGGTTGCGCTCT  
GGTGATATTGCTTATTATGACAATGATGGCCATTTTATATTGTGGACAGGCTGAAGTCATTAATTAATATAAAGGTT  
ATCAGGTTGCACCTGCTGAAATTGAGGGAATACTCTTACAACATCCGTATATTGTTGATGCCGGCGTTACTGGTATACC  
GGATGAAGCCGCGGGCGAGCTTCCAGCTGCAGGTGTTGTAGTACAGACTGGAAAATATCTAAACGAACAAATCGTACAA  
AATTTTGTTCAGTCAAGTTTCAACAGCCAAATGGCTACGTGGTGGGTGAAATTTTGGATGAAATCCCAAGGAT  
CAACTGGAATAATTGACAGAAAAGTGTAAAGACAAATGTTGAAAAACACAAATCTAAGCTG

## FIGURE 45

DPMEDKNILYGPEPFYPLADGTAGEQMFYALSRYADISGCIALTNAHTKENVLYEEFLKL  
SCRLAESFKKYGLKQNDTIAVCSENGLOFFLPLIASLYLGI AAPVSDKYIERELIHSIG  
IVKPRIIFCSKNTFOKVLNVKSKLKYVETIIILDLNEDLGGYQCLNNFISQNSDINLDVK  
KFKPNSFNRDDQVALVMFSSGTTGVSKGVMLTHKNIVARFSHCKDPTFGNAINPTTAILT  
VIPFHHGFGMMTTLGYFTCGFRVALMHTFEEKLFLOSLQDYKVESTLLVPTLMAFFAKSA  
LVEKYDLSHLKEIASGGAPLSKEIGEMVKKRFKLNFRQGYGLTETTSAVLITPDTDVRP  
GSTGKIVPFHAVKVVDPTTGKILGPNETGELYFKGDMIMKSYNNNEATKAIINKDGWLR  
SGDIAYYDNDGHFYIVDRKSLIKYQYQVAPAEIEGILLQHPYIVDAGVTGIPDEAAGE  
LPAAGVVVQTGKYLNEQIVQNFVSSQVSTAKWLRGGVKFLDEIPKGSTGKIDRKVLRQMF  
EKHKSKL

## FIGURE 46

DPMMKREKNVIYGPELHPLEDLTAGEMLFRALRKHSHLPQALVDVVGDESLSYKEFFEA  
TVLLAQSLHNCGYKMNDVVSICAENNTFFIPVIAAWYIGMIVAPVNESYIPDELCKVMG  
ISKPQIVFTTKNILNKVLEVQSRTNFIKRIIILDTVENIHGCESLPNFISRYSDGNIANF  
KPLHFDPEQVAAILCSSGTTGLPKGVMQTHQNICVRLIHALDPRAGTQLIPGVTVLVYL  
PFFHAFGFSITLGYFMVGLRVIMFRRFDQEAFLKAIQDYEVRSVINVPSVILFLSKSPLV  
DKYDLSSLRELCCGAAPLAKEVAEVAAKRLNLPGIRCGFGLTESTSANIHSRDEFKSGS  
LGRVTPLMAAKIADRETGKALGPNQVGELCIKGPVSKGYVNNVEATKEAIDDDGWLHSG  
DFGYDEDEHFYVVDYKELIKYKGSQVAPAELEEILLKNPCIRDVAVVGIPDLEAGELP  
SAFVVKQPGKEITAKEVYDYLAEVSHTKYLRGGVRFVDSIPRNVTKITRKELLKQLE  
KAGG

FIGURE 47

GGATCCCATGATGAAGCGAGAGAAAAATGTTATATATGGACCCGAACCCCTACACCCCTT  
GGAAGACTTAACAGCTGGAGAAATGCTCTTCGGTGCCTTCGAAAACATTCTCATTACC  
GCAGGCTTTAGTAGATGTGGTTGGCGACGAATCGCTTTCCTATAAAGAGTTTTTGAAGC  
GACAGTCCTCCTAGCGCAAAGTCTCCACAATTGTGGATACAAGATGAATGATGTAGTGTC  
GATCTGCGCCGAGAATAATAACAAGATTTTTTATTCCCGTTATTGCAGCTTGGTATATTGG  
TATGATTGTAGCACCTGTTAATGAAAGTTACATCCCAGATGAACTCTGTAAGGTGATGGG  
TATATCGAAACCACAAATAGTTTTTACGACAAAGAACATTTTAAATAAGGTATTGGAGGT  
ACAGAGCAGAACTAATTTCAATAAAAGGATCATCATACTTGATACTGTAGAAAACATACA  
CGGTTGTGAAAGTCTTCCCAATTTTATTCTCGTTATTTCGGATGGAAATATTGCCAATT  
CAAACCTTTACATTTTCGATCCTGTTGAGCAAGTGGCAGCTATCTTATGTTTCGTGAGGCAC  
TACTGGATTACCGAAAGGTGTAATGCAAACTCACCAAAATATTGTGTCCGACTTATACA  
TGCTTTAGACCCAGGGCAGGAACGCAACTTATTCCTGGTGTGACAGTCTTAGTATATCT  
GCCTTTTTTCCATGCTTTTGGGTTCTCTATAACCTTGGGATACTTCATGGTGGGTCTTCG  
TGTTATCATGTTGAGACGATTTGATCAAGAAGCATTTCATAAAGCTATTCAGGATTATGA  
AGTTGCAAGTGAATTAACGTTCCATCAGTAATATTGTTCTTATCGAAAAGTCCTTTGGT  
TGACAAATACGATTTATCAAGTTTAAGGGAATTGTGTTGCGGTGCGGCACCATTAGCAAA  
AGAAGTTGCTGAGGTTGCAGCAAAACGATTAACTTGCCAGGAATTCGCTGTGGATTGG  
TTTGACAGAATCTACTTCAGCTAATATACACAGTCTTAGGGATGAATTTAAATCAGGATC  
ACTTGAAGAGTTACTCCTTTAATGGCAGCTAAAATAGCAGATAGGGAACTGGTAAAGC  
ATTGGGACCAATCAAGTTGGTGAATTATGCATTAAAGGTCCCATGGTATCGAAAGGTTA  
CGTGAACAATGTAGAAGCTACCAAGAAGCTATTGATGATGATGGTTGGCTTCACTCTGG  
AGACTTTGGATACTATGATGAGGATGAGCATTCTATGTGGTGGACCGTTACAAGGAATT  
GATTAATATAAGGGCTCTCAGGTAGCACCTGCAGAACTAGAAGAGATTTTATTGAAAAA  
TCCATGTATCAGAGATGTTGCTGTGGTTGGTATTCCTGATCTAGAAGCTGGAGAAGTACC  
ATCTGCGTTTGTGGTTAAACAGCCCGAAAGGAGATTACAGCTAAAGAAGTGACGATTA  
TCTTGCCGAGAGGGTCTCCCATACAAAGTATTTGCGTGGAGGGGTCGATTGCTTGATAG  
CATACCAAGGAATGTTACAGGTAATAATTACAAGAAAGGAATTCTGAAGCAGTTGCTGGA  
GAAGGCGGGAGGT

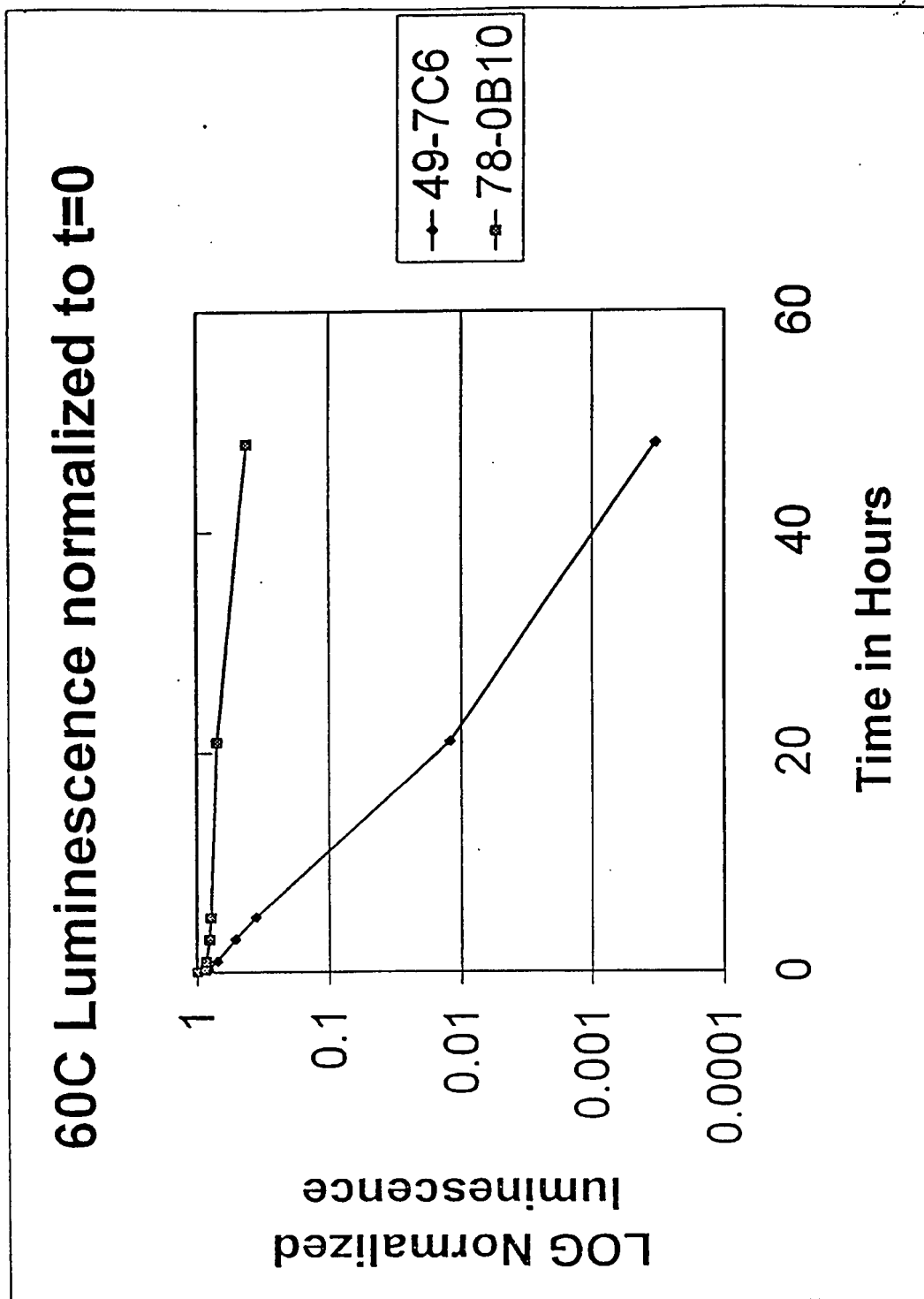


Fig 48

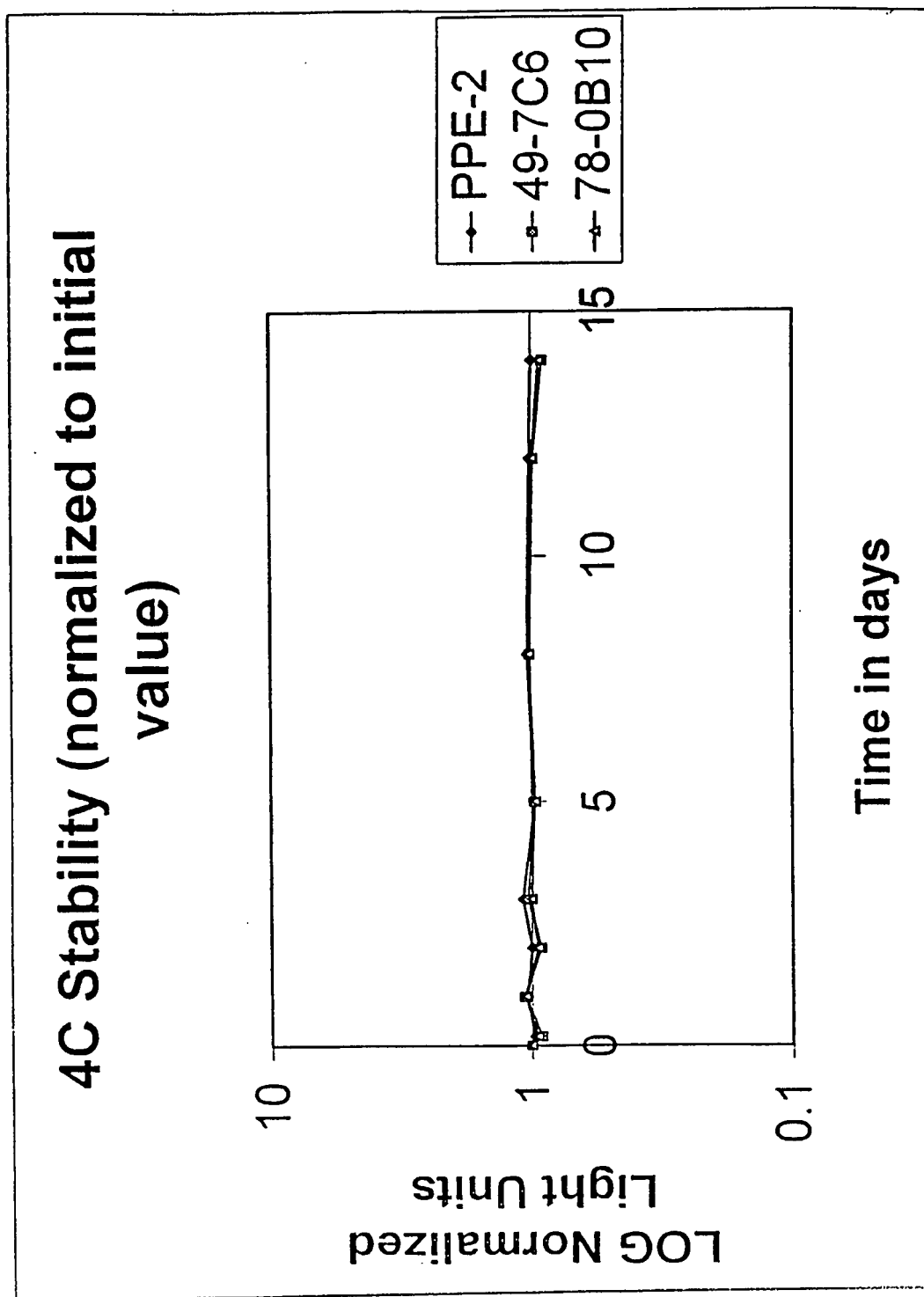


Figure 49



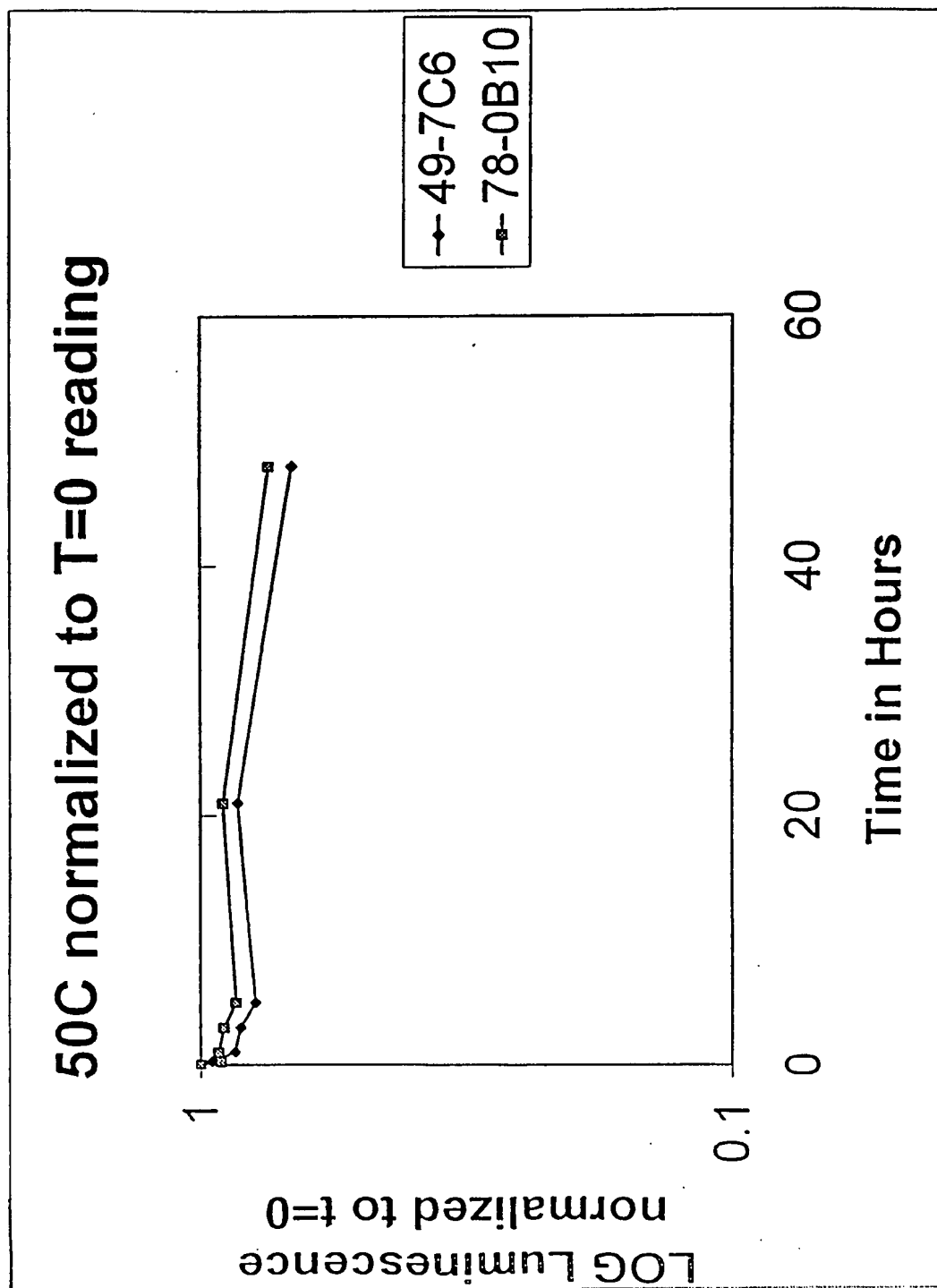


Figure 50

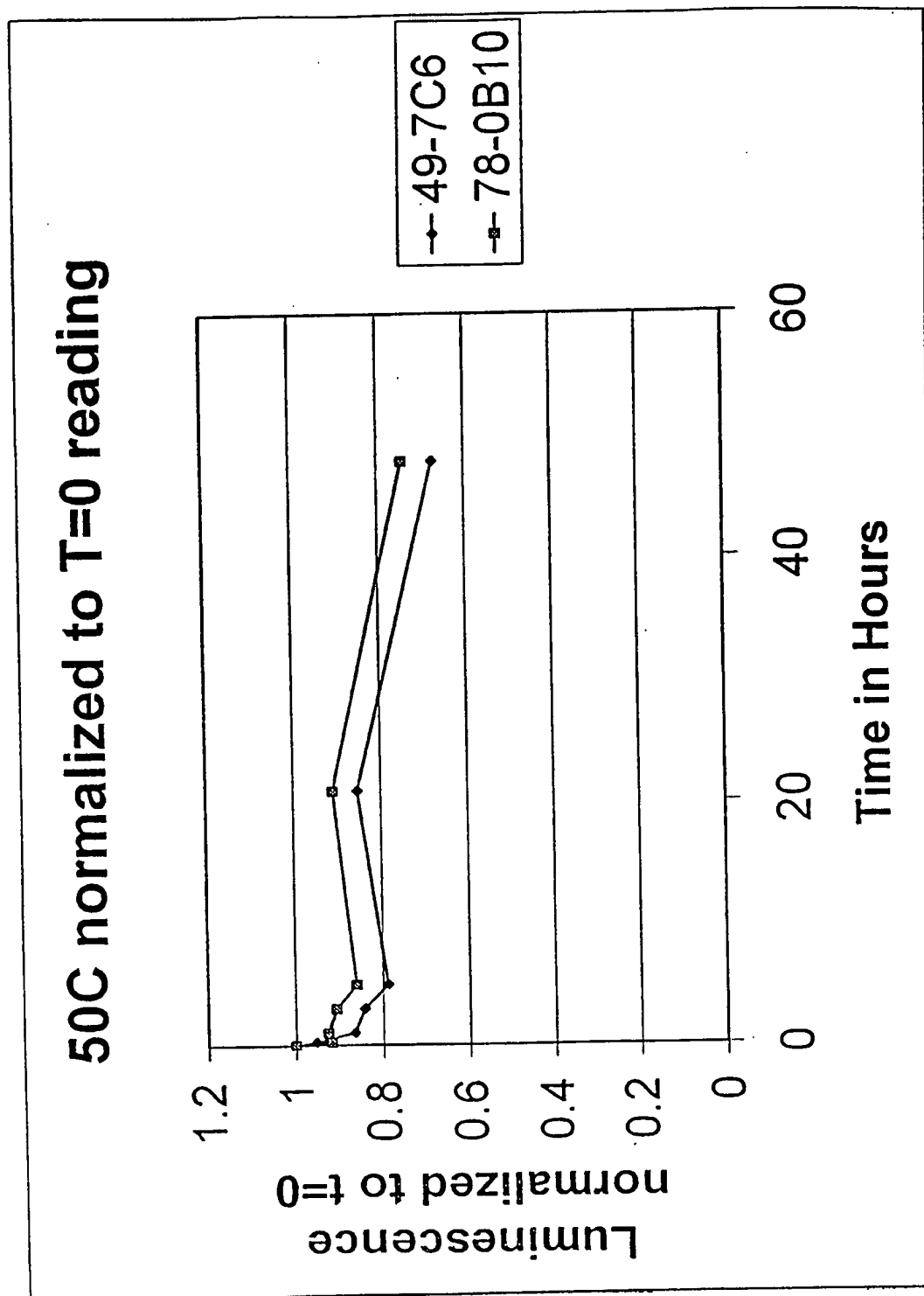


Figure 51

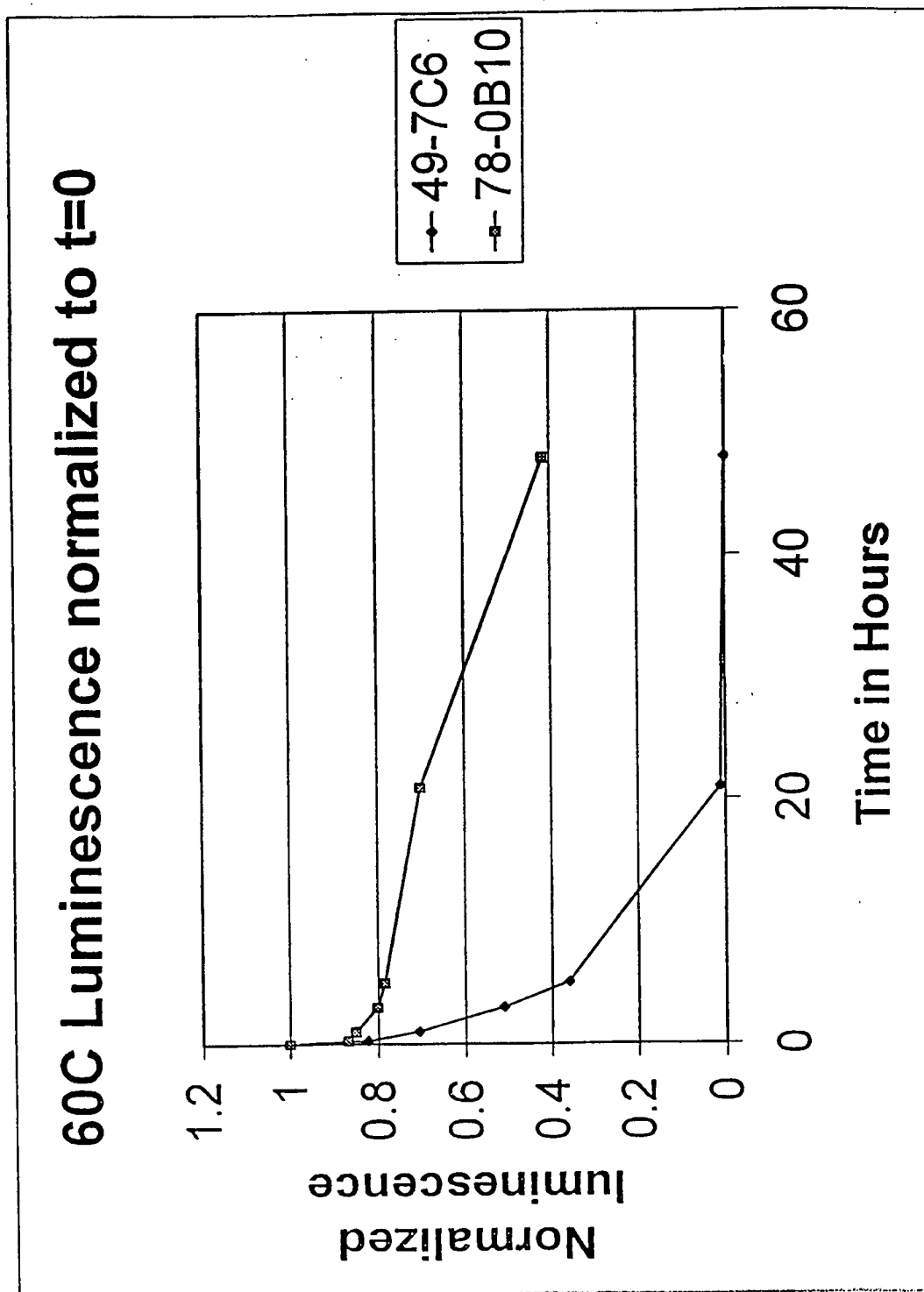


Figure 52

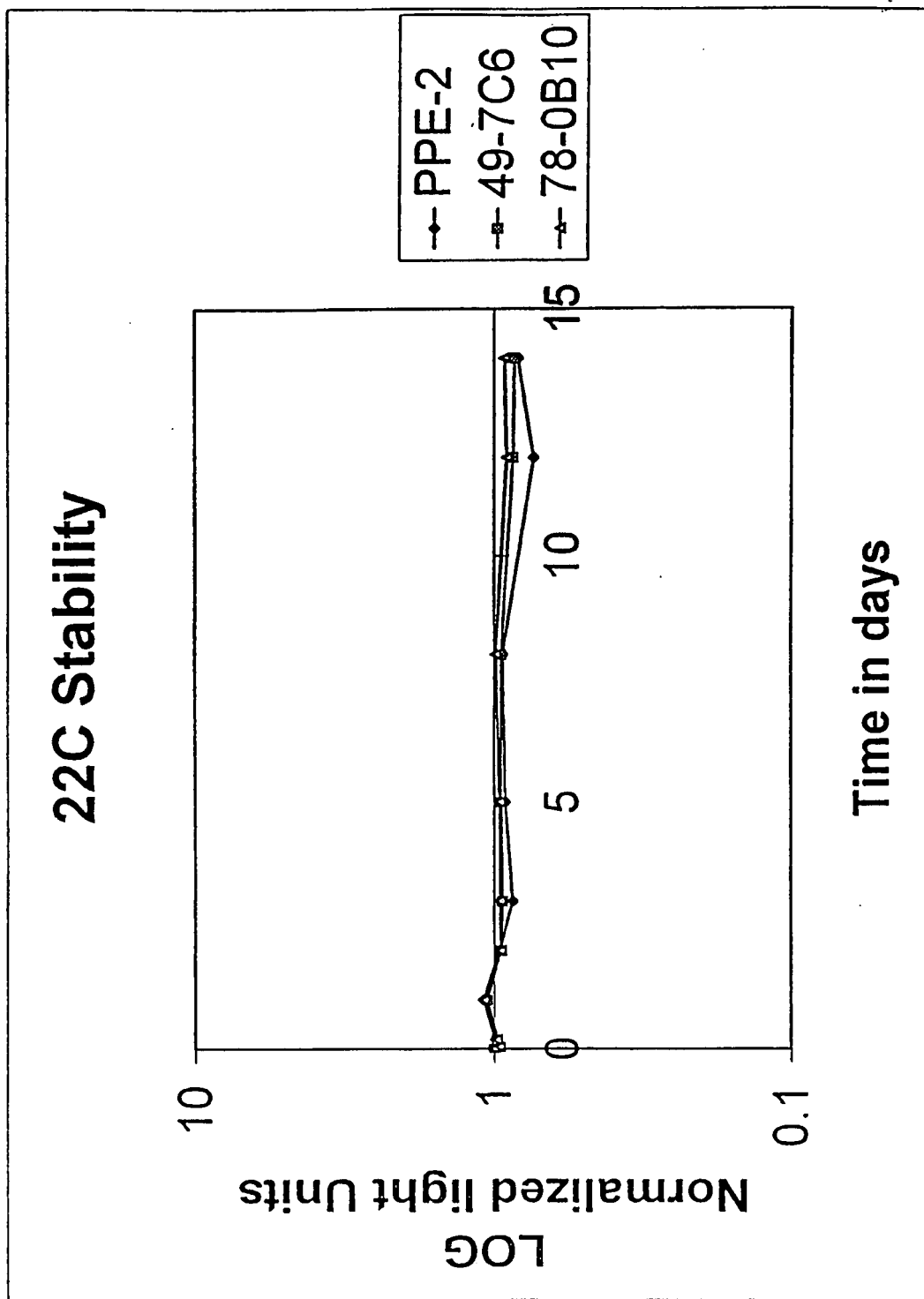


Figure 53

## Hierarchy of Polyadenylation Site Usage by Bovine Papillomavirus in Transformed Mouse Cells

ELLEN M. ANDREWS AND DANIEL DIMAIO\*

*Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510-8005*

Received 4 August 1993/Accepted 16 September 1993

**The great majority of viral mRNAs in mouse C127 cells transformed by bovine papillomavirus type 1 (BPV) have a common 3' end at the early polyadenylation site which is 23 nucleotides (nt) downstream of a canonical poly(A) consensus signal. Twenty percent of BPV mRNA from productively infected cells bypasses the early polyadenylation site and uses the late polyadenylation site approximately 3,000 nt downstream. To inactivate the BPV early polyadenylation site, the early poly(A) consensus signal was mutated from AAUAAA to UGUAAA. Surprisingly, this mutation did not result in significant read-through expression of downstream RNA. Rather, RNA mapping and cDNA cloning experiments demonstrate that virtually all of the mutant RNA is cleaved and polyadenylated at heterogeneous sites approximately 100 nt upstream of the wild-type early polyadenylation site. In addition, cells transformed by wild-type BPV harbor a small population of mRNAs with 3' ends located in this upstream region. These experiments demonstrate that inactivation of the major poly(A) signal induces preferential use of otherwise very minor upstream poly(A) sites. Mutational analysis suggests that polyadenylation at the minor sites is controlled, at least in part, by UAUUA, an unusual variant of the poly(A) consensus signal approximately 25 nt upstream of the minor polyadenylation sites. These experiments indicate that inactivation of the major early polyadenylation signal is not sufficient to induce expression of the BPV late genes in transformed mouse cells.**

Eukaryotic RNA polymerase II transcription units are typically transcribed past the mature mRNA 3' end. These transcripts are then cleaved and a poly(A) tract of 200 to 300 nucleotides (nt) is added to generate the 3' end of the mature mRNA (for reviews, see references 30 and 37). Eighty to ninety percent of animal cell mRNAs contain the sequence AAUAAA 10 to 30 nt upstream of the poly(A) tail. Another 10% have the variant AUUAAA; other variants are rare (38). These consensus sequences have been shown to be required for efficient and accurate cleavage and polyadenylation both *in vivo* and *in vitro* (17, 27, 29). Generally, when this sequence is mutated, polyadenylation occurs at a downstream site, often with reduced efficiency (17). The region upstream and downstream of the AAUAAA consensus signal, including GU-rich downstream sequences, has also been identified as playing a role in the cleavage and polyadenylation of some transcripts (30, 37).

Bovine papillomavirus type 1 (BPV) induces fibropapillomas in cattle and transforms a number of cultured rodent fibroblast cell lines to tumorigenicity. The papillomaviruses are unable to propagate in such transformed cells, in part because the early polyadenylation site used by essentially all BPV transcripts in transformed cells is located between the transcriptional promoters and L1 and L2, the two genes which encode the virion proteins (Fig. 1A) (16, 23, 39). Similarly, in BPV-induced skin fibropapillomas, usage of this early polyadenylation site precludes expression of the capsid protein genes in transformed dermal fibroblasts and presumably in the basal keratinocytes as well (4, 5, 35). In terminally differentiating keratinocytes which express the capsid proteins and produce virus, about 20% of the viral mRNA reads through the early polyadenylation site and is instead polyadenylated approximately 3,000 nt downstream at the late polyadenylation site (5). Thus, regulation of

polyadenylation at the early site appears to be crucial for viral late gene expression.

To study signals that control polyadenylation in BPV-transformed mouse C127 cells, we mutated the early poly(A) consensus signal AAUAAA, located 23 nt upstream of the early poly(A) site. It was expected that mutant transcripts would now bypass the early poly(A) site and that late region sequences would be included in stable RNA. RNA mapping experiments instead demonstrated that mutant transcripts were polyadenylated at heterogeneous sites approximately 100 nt upstream of the early polyadenylation site used in cells transformed by wild-type BPV. Evidence is presented which suggests that an unusual variant of the poly(A) consensus sequence, UAUUA, plays a role in the regulation of polyadenylation at the upstream polyadenylation sites.

**Construction and preliminary characterization of the poly(A) consensus mutant.** To disrupt polyadenylation at the BPV major early polyadenylation site at nt 4203, oligonucleotide-directed mutagenesis was used to mutate the poly(A) consensus signal at nt 4180 from AAUAAA to UGUAAA (Fig. 1B), thereby creating a new *PvuII* cleavage site (23a). The resulting mutation on a *BstXI*-to-*Sall* fragment was reconstructed into the full-length wild-type BPV genome (clone pBPV-142-6 [33]) to generate mutant pBPV-EPA1. Nucleotide sequence analysis of the fragment replaced in generating pBPV-EPA1 (nt 3849 and 4450) demonstrated that no extraneous mutations were introduced during mutagenesis.

The ability of three isolates of pBPV-EPA1 to transform C127 cells was assayed by determining the efficiency of focus formation after *Bam*HI digestion to release the viral DNA from the plasmid vector and transfection as described previously (14). All three mutant isolates transformed cells with approximately the same efficiency as wild-type BPV DNA (data not shown). Cell lines were derived from pools of foci induced by pBPV-EPA1 (EPA1p) and by wild-type BPV DNA (142-6p). ID13 cells, a C127 cell line transformed by infection with BPV, were used as an additional wild-type control.

\* Corresponding author.

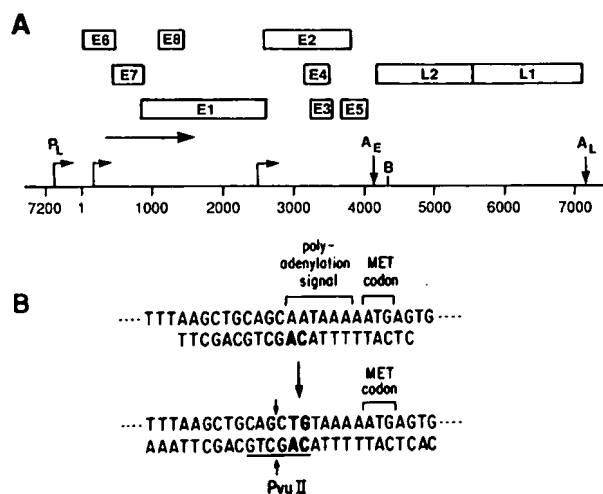


FIG. 1. The BPV genome and design of the EPA1 mutation. (A) The BPV genome linearized at the 3' end of the late transcription unit is shown. Open boxes indicate translational open reading frames. The long horizontal arrow indicates the direction of transcription. The short horizontal arrows indicate the positions of major promoters. The late promoter is designated P<sub>L</sub>. A<sub>E</sub> and A<sub>L</sub> denote the early and late polyadenylation sites, respectively. B indicates the position of the unique *Bam*HI site. Nucleotide numbers are shown at the bottom. (B) The EPA1 mutation. The top line shows the wild-type BPV DNA sequence around the early polyadenylation signal. The sequence of the mutagenic oligonucleotide L1 is shown directly below it, with the base substitutions shown in boldface. The template was the small *Bam*HI to *Eco*RI fragment of BPV-1 DNA cloned in M13mp8 (13). The sequence at the bottom shows the mutation with the new *Pvu*II site indicated. The open reading frame L2 initiation codon is designated the MET codon.

Southern blot analysis of viral DNA from transformed cells demonstrated that the mutant viral DNA was maintained in transformed cells as a multicopy plasmid without gross rearrangement and with restoration of the *Bam*HI site used to excise the viral DNA from the plasmid vector (data not shown).

**Mapping the 3' end of the mutant mRNA.** The mutation in pBPV-EPA1 was designed to eliminate polyadenylation at the wild-type early polyadenylation site immediately upstream of the late open reading frames. Extensive Northern (RNA) blot analysis and RNA protection experiments failed to detect significant amounts of RNA extending past the polyadenylation site into the late region, but these experiments did not exclude the presence of low levels of read-through RNA (data not shown). There was severalfold more stable viral RNA in cells transformed by wild-type BPV than in those transformed by the polyadenylation site mutant (Fig. 2).

RNAse protection experiments were performed to map the 3' ends of the mutant transcripts. ID13 and EPA1p RNAs were assayed for protection of an antisense EPA1 RNA probe spanning the early polyadenylation site at nt 4203 (Fig. 2, left panel). The size of the fragment protected by RNA from ID13 cells indicates that, as expected, the wild-type viral RNA extends past nt 4180, the site of the mutation in the probe (lane c). In contrast, EPA1p RNA protected several fragments approximately 100 nt shorter than those protected by wild-type RNA, suggesting that the mutant RNA is polyadenylated upstream of the normal position (lanes a and b). The difference in the pattern of protected bands between the two EPA1p

lanes, a and b, is due to the different cleavage specificities of the two RNases used in these reactions. The same result was obtained with oligo(dT)-selected EPA1p RNA (data not shown), indicating that these shorter species are polyadenylated, a conclusion confirmed by cDNA cloning (see below). There was no evidence of significant polyadenylation of EPA1p RNA at the usual position, nor were prominent shorter novel bands protected in the ID13 sample. RNA from two additional cell lines generated with the original isolate of the mutant and two additional cell lines generated with independent isolates of the mutant showed the protection pattern characteristic of the mutant (data not shown). These results suggested that sequences downstream of nt 4100 were absent from mutant RNA, an interpretation supported by the results of protection experiments with additional antisense probes and the results of Northern blot hybridization experiments with oligonucleotide probes (data not shown). These results are interpreted in the right panel of Fig. 2.

**cDNA cloning and sequencing.** The results presented above suggest that new heterogeneous polyadenylation sites near nt 4100 are utilized in EPA1p RNA. To confirm this interpretation, the 3' ends of both wild-type and mutant RNAs were cloned and sequenced. Oligo(dT)-selected (3) 142-6p and EPA1p RNAs were reverse transcribed with oligo(dT) as primer and the reagents and protocol of a cDNA synthesis kit (Amersham). The resulting first-strand cDNAs were amplified by the polymerase chain reaction (PCR) method with the primers diagrammed in Fig. 3A (18, 32, 34). To specifically amplify BPV sequences, the upstream PCR primer PCR5 corresponded to BPV nt 3998 to 4031. To selectively amplify polyadenylated molecules, the downstream PCR primer PCRT was 5' d(GGGGATCCT<sub>25</sub>) 3', which hybridized to any product containing a poly(A) tract. Annealing was carried out at 25°C, because PCRT has a calculated *T<sub>m</sub>* of 38.9°C in PCR buffer conditions (32). The products of each amplification reaction were cloned into pUC18, and colonies containing an insert were identified by colony hybridization (22) with an oligonucleotide probe PCR1 complementary to a region (nt 4063 to 4089) between the upstream primer and the proposed 3' end of mutant RNA.

The results of sequence analysis of the cDNA clones are summarized in Fig. 3B. Sites of polyadenylation were identified as junctions between BPV DNA sequence and tracts of poly(A). Six of the 11 clones derived from cells transformed by wild-type BPV were polyadenylated after nt 4203, the previously described early polyadenylation site (39), thus validating this strategy of identifying polyadenylation sites. In contrast, none of the clones derived from mutant RNA displayed the wild-type polyadenylation site. Instead, seven of the nine EPA1p clones contain a stretch of poly(A) immediately after BPV nt 4107, and the other two clones contain poly(A) after nt 4101 and 4092. These results are consistent with the RNAse protection and Northern blot results which indicate the existence of heterogeneous 3' ends near nt 4100 in mutant RNA and demonstrate that these new 3' ends are in fact new sites of polyadenylation. Interestingly, the anomalous clones (almost half) derived from wild-type RNA showed polyadenylation at heterogeneous sites similar to those found with the mutant RNA. These results indicate that there is a population of mRNAs with heterogeneous polyadenylation sites around nt 4100 in cells transformed by wild-type BPV. The preferential amplification of shorter PCR products may explain the relatively frequent isolation of these shorter cDNAs from cells transformed by wild-type BPV. We have occasionally observed faint bands in protection experiments with ID13 RNA which

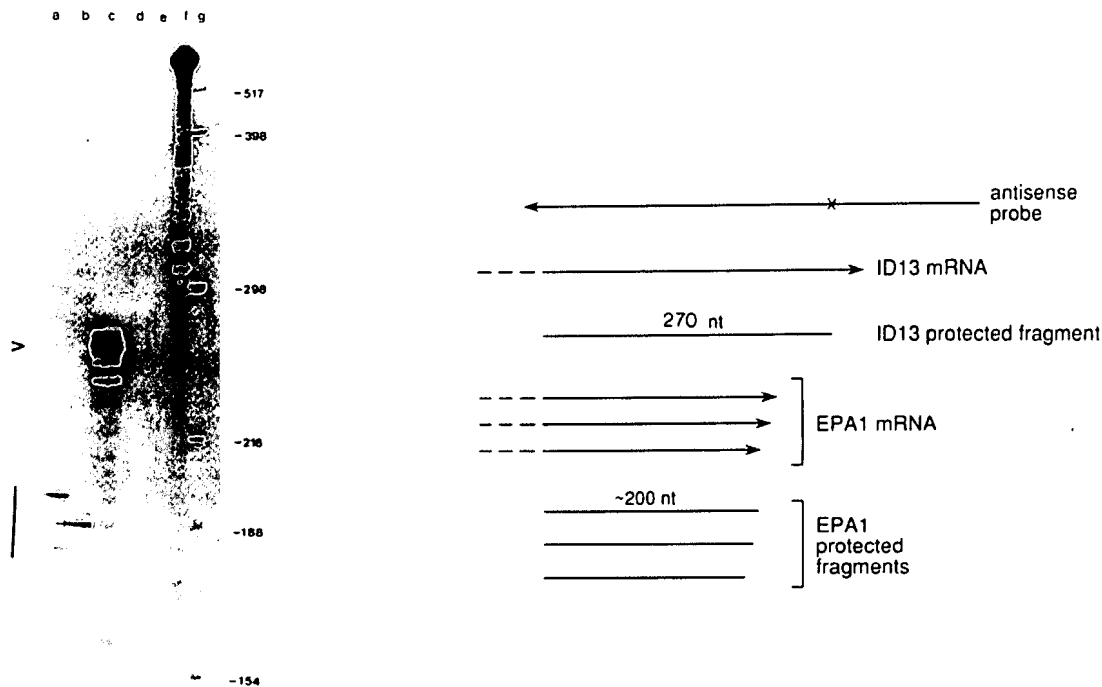


FIG. 2. RNase protection analysis of viral early region RNA in transformed cells. (Left panel) Ten micrograms of total cellular RNA (9, 21) from EPA1p (lanes a and b), ID13 (lane c), and C127 (lane d) cells was hybridized to an antisense EPA1 RNA probe (28) complementary to BPV nt 3912 to 4450, spanning the position of the normal early polyadenylation site but containing the mutation at the polyadenylation signal. Hybrids were digested with either 2  $\mu$ g of RNase T1 per ml (lane a), 40  $\mu$ g of RNase A per ml (lane b), or a mixture of both RNases (lanes c to e), and protected fragments were detected by autoradiography after electrophoresis through a 4% polyacrylamide–50% urea gel. The sample in lane e was the probe digested after mock hybridization; the sample in lane f is undigested probe. The nucleotide lengths of size markers in lane g are indicated. The arrowhead indicates the predicted position of a 270-nt fragment extending from the 3' end of the probe to the site of the mutation, which is generated by cleavage at the mismatch between the wild-type RNA and the mutation in the probe. The vertical line on the left indicates the small cluster of bands protected by mutant RNA. (Right panel) Schematic representation of the probe, protected fragments generated by RNase digestion, and the deduced structure of viral RNA species. Arrows indicate the direction of transcription, with the arrowheads representing the 3' end of each transcript. The X indicates the position of the mutation in the probe.

are consistent with minor sites of polyadenylation at these upstream positions (data not shown).

**Identification of a signal controlling polyadenylation at the upstream sites.** There is no poly(A) consensus sequence or previously described functional variant within 100 bp upstream of nt 4100. However, the sequence UAUUA is present at nt 4073, approximately 30 nt 5' to the poly(A) sites in EPA1p RNA (Fig. 3B). It is the closest match to the consensus sequence in the region, and it appears to be in the appropriate position to specify cleavage at the sites detected in mutant RNA. To test the role of this sequence in specifying polyadenylation in the absence of the wild-type signal, it was mutated from UAUUA to GAUAUC by using the mutagenic primer 5' d(AACTTCATAC AGGATATCAA ACAAATCA)3', corresponding to BPV sequence from nt 4063 to 4090, and single-stranded EPA1 DNA as a template. The resulting mutant, pBPV-EPA2 (see Fig. 5) therefore contained both the original mutation at the poly(A) consensus signal and the new mutations in the putative variant signal. This mutant transformed C127 cells with approximately wild-type efficiency, and RNA from a pooled cell line transformed by EPA2 DNA was mapped by using RNase protection and an antisense EPA2 probe (Fig. 4). RNA from ID13 cells protected the fragment sizes predicted if cleavage occurred at the sites of mismatch

between wild-type RNA and the probe (which contains mutations at nt 4073, 4078, 4180, and 4181) (lane b). EPA2 RNA protected two major size classes of fragments (lane a). One was a set of probe fragments approximately 190 to 200 nt long, corresponding to polyadenylation near nt 4100 as in EPA1p RNA. These protected fragments comigrate with the fragments protected by EPA1 RNA (data not shown) and are the size predicted if polyadenylation occurred at nt 4107, the mutant site mapped by cDNA cloning. In addition, EPA2p RNA protected several longer fragments corresponding to heterogeneous RNA 3' ends between nt 4200 and 4450. The EPA2 mutation thus reduced the efficiency with which the upstream polyadenylation sites are used, but it did not appear to affect the position of poly(A) addition for those transcripts that are successfully polyadenylated in this region. These results indicate that the UAUUA plays a role in specifying the new upstream sites of polyadenylation in EPA1p RNA.

**Discussion.** These experiments were designed to study polyadenylation site usage in BPV-transformed mouse cells. A point mutation in the poly(A) consensus signal disrupted polyadenylation at that site both in vivo, as demonstrated here, and in an in vitro polyadenylation system (24). RNase protection, Northern blotting, and cDNA cloning and sequencing established that stable mutant transcripts utilized heteroge-

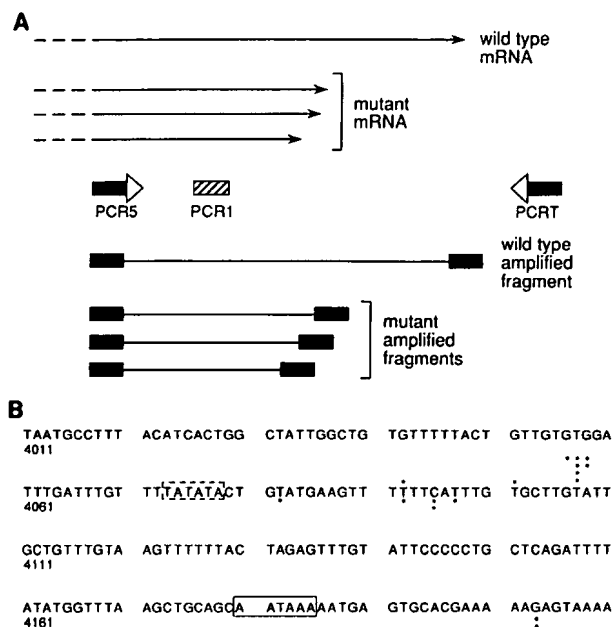


FIG. 3. (A) PCR-based strategy to clone the 3' ends of viral RNA. The top portion of the panel shows the deduced structure of the 3' ends of the major viral RNAs. The upstream primer, PCR5, is complementary to all known wild-type and mutant viral early RNAs. The downstream primer, PCRT, consists of oligo(dT) and a cloning site but no BPV-specific sequences. After amplification with cDNA reverse transcribed from polyadenylated RNA as a template, BPV cDNAs were cloned into pUC18, identified by hybridization to PCR1, and sequenced. (B) cDNA clone sequences. The sense strand BPV sequence from nt 4011 to 4210 is shown. The normal poly(A) consensus signal is enclosed in the solid line, and the putative upstream poly(A) signal is enclosed in the dashed line. Each dot indicates the position of a junction between BPV DNA and the poly(A) tract in a cDNA clone. Clones derived by amplification of wild-type RNA are represented by dots below the sequence, and those derived from mutant RNA are represented by dots above the sequence.

neous polyadenylation sites approximately 100 nt upstream of the wild-type polyadenylation site. The results of the cDNA cloning also demonstrated that some wild-type transcripts have heterogeneous 3' ends in the region used by the mutant RNAs, indicating that this is a minor polyadenylation site in cells transformed by wild-type BPV. In a related system, Doniger et al. (15) found usage of upstream polyadenylation sites by human papillomavirus type 16 transcripts in an immortalized human exocervical epithelial cell line harboring a human papillomavirus type 16 genome with an extensive deletion immediately downstream of a wild-type early poly(A) signal. The 3' ends of viral RNA from these cells mapped to both the normal site and to a heterogeneous region 400 to 500 nt upstream of that site.

The results described here suggest a hierarchy of polyadenylation site usage in BPV-transformed cells, as is summarized in Fig. 5. Wild-type BPV mRNA is polyadenylated at the major polyadenylation site at nt 4203, with a small fraction of transcripts being polyadenylated at minor upstream sites around nt 4100. When the major signal is disrupted (as in EPA1), the sites around nt 4100 become the predominant sites of polyadenylation. When the major signal is inactivated and the minor signal is partially disrupted (as in EPA2), both the

upstream sites and new downstream sites between nt 4200 and 4450 are used. Additional experiments have shown that polyadenylation occurs exclusively at these downstream sites when the major signal is inactivated and the upstream polyadenylation region is deleted (2). There are several potential polyadenylation signals in this downstream region, including a sequence at nt 4304 that deviates by 1 nt from the consensus polyadenylation signal. In addition, Burnett et al. (8) observed polyadenylation near nt 4450 in RNA from cells transformed by a spontaneous BPV-1 deletion mutant lacking the major poly(A) site and surrounding sequences. One can speculate that the function of the multiple potential early polyadenylation sites in BPV is to ensure that late genes are not expressed under inappropriate conditions, for example in transformed dermal fibroblasts or basal epidermal keratinocytes.

Polyadenylation site selection appears to be a complex process that takes into account both the relative strengths of potential sites and their positions relative to one another (12, 20). Moreover, the representation of polyadenylation sites in stable RNA reflects a number of factors in addition to polyadenylation site selection, including the stability of various RNA species. The results of the RNase protection experiments reported here indicate that the upstream polyadenylation sites are used far more abundantly by the early polyadenylation signal mutant than by the wild type. However, it is also clear that there is less total viral RNA in cells transformed by the mutant. It is possible that processing at the upstream sites remains relatively inefficient even with the mutant polyadenylation signal, resulting in the synthesis of a rapidly degraded pool of unprocessed RNA extending into the late region. In fact, Furth and Baker (19) have described a sequence element in the BPV late region which prevents the accumulation of stable viral RNA in transformed cells.

The closest match to a poly(A) consensus signal in the vicinity of the minor upstream polyadenylation sites is UAUUAU, approximately 25 nt upstream of the new RNA 3' ends. RNA from cells containing mutations of both the original poly(A) signal and this putative upstream signal contains heterogeneous 3' ends at both the upstream sites and at additional positions downstream of the normal site. This result suggests that the UAUUAU plays a role in directing polyadenylation at the upstream polyadenylation sites and that the mutation did not fully disrupt the function of the UAUUAU sequence. We are not aware of a precedent for UAUUAU acting as a poly(A) signal in mammalian cells, although it can direct mRNA 3' end formation and polyadenylation in *Saccharomyces cerevisiae* (31). However, we note that the region around the upstream cleavage sites contains numerous oligo(dT) tracks and GT dinucleotides, sequence motifs found near some bona fide mammalian poly(A) signals.

The wild-type poly(A) consensus signal appears to suppress utilization of the variant signal located approximately 100 nt upstream. Such suppression may be rather general. Connelly and Manley (10) studied the simian virus 40 early polyadenylation region, which contains two closely spaced AAUAAA signals. In the wild-type situation, only the 3' site is efficiently utilized. However, if this preferred site was inactivated by mutation, increased usage of the 5' site was observed. In addition, Denome and Cole (11) showed that addition of tandemly arranged polyadenylation signals decreased usage of the upstream site. These findings imply that genomes may contain numerous potential sites of polyadenylation whose activity is suppressed by the relatively close apposition of another polyadenylation signal, which perhaps competes more efficiently for a limiting polyadenylation factor. Therefore, alternative polyadenylation, which is a well-documented con-



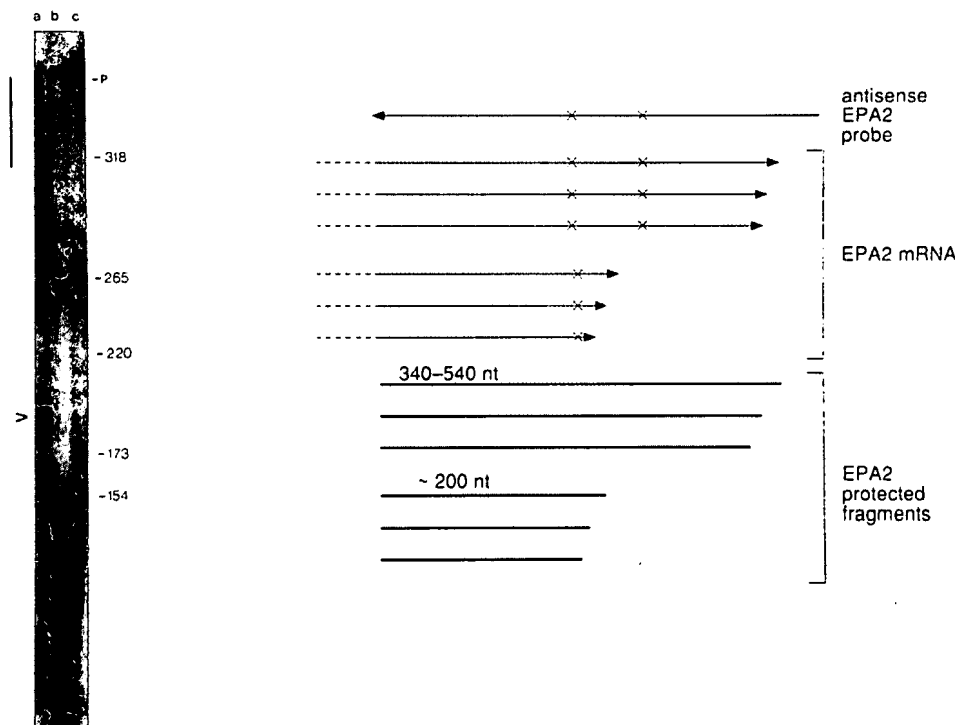


FIG. 4. Evidence that UAUUA at nt 4073 plays a role in poly(A) site selection. (Left panel) Radiolabelled antisense RNA probe extending from BPV nt 4450 to 3912 was transcribed in vitro from EPA2, hybridized to 10  $\mu$ g of cellular RNA isolated from EPA2p (lane a), ID13 (lane b), or C127 (lane c) cells, and digested with a mixture of RNases A and T1. Protected fragments were subjected to polyacrylamide gel electrophoresis and detected by autoradiography. The arrowhead on the left indicates the position of approximately 190- to 200-nt probe fragments extending from the 3' end of the probe to the upstream sites of polyadenylation around BPV nt 4100. The vertical line on the left indicates the position of the larger fragments also protected by mutant RNA. The approximately 265-base fragment in lane b appears to be derived from partially digested hybrids. The lengths (in nucleotides) of coelectrophoresed size markers are shown. P indicates the position of undigested probe (538 nt). (Right panel) Schematic representation of the antisense probe, sizes of the protected fragments, and deduced structures of viral RNA species. The X's show the positions of the mutations at the upstream and downstream polyadenylation signals in the probe and in RNA isolated from cells transformed by the double mutant.

trol point for regulating gene expression (25), may result in some cases from inactivation of a preferred polyadenylation signal rather than by direct activation of a suboptimal one.

The mechanism by which BPV prevents expression of viral

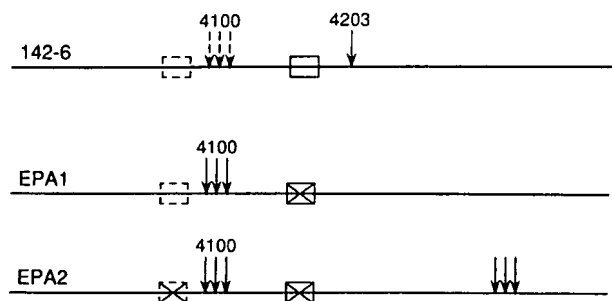


FIG. 5. Usage of early region polyadenylation sites. The horizontal lines represent the region of the BPV genome around the early polyadenylation site for wild-type BPV DNA (142-6) and the indicated mutants. Transcription proceeds from left to right. The unbroken box represents the normal early polyadenylation consensus signal at nt 4180, and the dashed boxes represent the putative upstream polyadenylation signal at nt 4073. The vertical arrows indicate the positions of poly(A) addition, and triple arrows indicate heterogeneous polyadenylation sites, with minor sites represented as dashed arrows. Boxes containing an X indicate a mutant polyadenylation signal.

late genes in transformed cells but allows their expression in differentiated keratinocytes is central to an understanding of papillomavirus biology. One level of restriction in transformed cells is clearly at the level of stable mRNA accumulation, because little or no BPV mRNA from the late region is present in cultured fibroblasts. Analysis of nascent RNA from ID13 cells indicates that at least 90% of BPV transcripts terminate between the early and late poly(A) sites and therefore never reach the late poly(A) signal (6). The mechanism(s) allowing production of late RNAs during natural infection may act primarily at the level of polyadenylation site selection, or it may act at some other steps in mRNA biogenesis, such as alterations in promoter usage, splicing patterns, or transcription termination, which secondarily affect cleavage and polyadenylation (for examples, see references 1, 7, 26, and 36). However, the results presented here indicate that specific suppression of the major early polyadenylation signal is unlikely to be the sole step in releasing the block to BPV late gene expression, because inhibition of polyadenylation at additional potential early sites must also occur. The mechanism involved in late gene expression must coordinately suppress cleavage and polyadenylation at multiple potential sites near the 3' end of the early region in some of the transcripts, while many transcripts are still polyadenylated at the early polyadenylation site. Regulation of BPV late gene expression is clearly a complex process and bypass of the early major

polyadenylation site is a necessary but not sufficient component of that process. The study of BPV transcriptional regulation promises to provide insights into not only papillomavirus biology but also the mechanisms of regulation of gene expression in general.

We thank S. Lee, S. Amara, and D. Ward for helpful discussions; S. Baserga and J. Brandsma for reviewing the manuscript; and J. Zulkowski for assistance in manuscript preparation.

This work was supported by a grant from the National Institutes of Health (CA37157).

#### REFERENCES

- Acheson, N. H. 1984. Kinetics and efficiency of polyadenylation of late polyomavirus nuclear RNA: generation of oligomeric polyadenylated RNAs and their processing into mRNA. *Mol. Cell. Biol.* 4:722-729.
- Andrews, E., and D. DiMaio. Unpublished observations.
- Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* 69:1408-1412.
- Baker, C. 1989. Bovine papillomavirus type 1 transcription, p. 91-112. In H. Pfister (ed.), *Papillomaviruses and human cancer*. CRC Press, Boca Raton, Fla.
- Baker, C., and P. Howley. 1987. Differential promoter utilization by the bovine papillomavirus in transformed cells and in productively infected wart tissues. *EMBO J.* 6:1027-1035.
- Baker, C., and J. S. Noe. 1989. Transcriptional termination between bovine papillomavirus type 1 (BPV-1) early and late polyadenylation sites blocks late transcription in BPV-1-transformed cells. *J. Virol.* 63:3529-3534.
- Brown, P. H., L. S. Tiley, and B. R. Cullen. 1991. Effect of RNA secondary structure on polyadenylation site selection. *Genes Dev.* 5:1277-1284.
- Burnett, S., J. Moreno-Lopez, and U. Pettersson. 1988. A novel spontaneous mutation of the bovine papillomavirus-1 genome. *Plasmid* 20:61-74.
- Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294-5301.
- Connelly, S., and J. L. Manley. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2:440-452.
- Denome, R. M., and C. N. Cole. 1988. Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. *Mol. Cell. Biol.* 8:4829-4839.
- DeZazzo, J. D., and M. J. Imperiale. 1989. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol. Cell. Biol.* 9:4951-4961.
- DiMaio, D. 1986. Nonsense mutation in open reading frame E2 of bovine papillomavirus DNA. *J. Virol.* 57:475-480.
- DiMaio, D., R. H. Treisman, and T. Maniatis. 1982. Bovine papilloma virus vector that propagates as a plasmid in both mouse and bacterial cells. *Proc. Natl. Acad. Sci. USA* 79:4030-4034.
- Doniger, J., C. D. Woodworth, and J. A. DiPaolo. 1990. Alternative HPV 16 early message termination sites. *UCLA Symp. Mol. Cell. Biol.* 124:519-522.
- Engel, L. W., C. A. Heilman, and P. M. Howley. 1983. Transcriptional organization of bovine papillomavirus type 1. *J. Virol.* 47:516-528.
- Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24:251-260.
- Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid amplification of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85:8998-9002.
- Furth, P. A., and C. C. Baker. 1991. An element in the bovine papillomavirus late 3' untranslated region reduces polyadenylated cytoplasmic RNA levels. *J. Virol.* 65:5806-5812.
- Galli, G., J. W. Guise, M. A. McDevitt, P. W. Tucker, and J. R. Nevins. 1987. Relative position and strengths of poly(A) sites as well as transcription termination are critical to membranes versus secreted  $\mu$ -chain expression during B-cell development. *Genes Dev.* 1:471-481.
- Glisin, V., R. Crkvenjakov, and C. Byus. 1973. Ribonucleic acid isolated by cesium chloride centrifugation. *Biochemistry* 13:2633-2637.
- Grundstein, M., and D. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNA's that contain a specific gene. *Proc. Natl. Acad. Sci. USA* 72:3961-3965.
- Heilman, C. A., L. Engel, D. R. Lowy, and P. M. Howley. 1982. Virus-specific transcription in bovine papillomavirus-transformed mouse cells. *Virology* 119:22-34.
- Kunkel, T. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* 82:488-492.
- Lee, S., and E. Andrews. Unpublished observations.
- Leff, S. E., M. G. Rosenfeld, and R. M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* 55:1091-1118.
- Logan, J., E. Falk-Pederson, J. E. Darnell, and T. Shenk. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse B<sup>mini</sup>-globin gene. *Proc. Natl. Acad. Sci. USA* 84:8306-8310.
- Manley, J. L., H. Yu, and L. Ryner. 1985. RNA sequence containing hexanucleotide AAUAAA directs efficient mRNA polyadenylation in vitro. *Mol. Cell. Biol.* 5:373-379.
- Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035-7056.
- Orkin, S., T. Cheng, S. Antonarakis, and H. Kazazian. 1985. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human  $\beta$ -globin gene. *EMBO J.* 4:453-456.
- Proudfoot, N. 1991. Poly(A) signals. *Cell* 64:671-674.
- Russo, P., W.-Z. Li, D. M. Hampsey, K. S. Zaret, and F. Sherman. 1991. Distinct cis-acting signals enhance 3' endpoint formation of *CYC1* mRNA in the yeast *Saccharomyces cerevisiae*. *EMBO J.* 10:563-571.
- Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Sarver, N., J. C. Byrne, and P. M. Howley. 1982. Transformation and replication in mouse cells of a bovine papillomavirus-pML2 plasmid vector that can be rescued in bacteria. *Proc. Natl. Acad. Sci. USA* 79:7147-7151.
- Scharf, S. J., G. T. Horn, and H. A. Erlich. 1986. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* 233:1076-1078.
- Stoler, M. H., and T. R. Broker. 1986. In situ hybridization detection of human papillomavirus DNAs and messenger RNAs in genital condylomas and a cervical carcinoma. *Hum. Pathol.* 17:1250-1258.
- Whitelaw, E., and N. Proudfoot. 1986.  $\alpha$ -Thalassemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human  $\alpha$ -2 globin gene. *EMBO J.* 5:2915-2922.
- Wickens, M. 1990. How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem. Sci.* 15:277-281.
- Wickens, M., and P. Stephenson. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutations prevent messenger RNA 3' end formation. *Science* 226:1045-1051.
- Yang, Y. C., H. Okayama, and P. M. Howley. 1985. Bovine papillomavirus contains multiple transforming genes. *Proc. Natl. Acad. Sci. USA* 82:1030-1034.



## DIRECTED EVOLUTION: CREATING BIOCATALYSTS FOR THE FUTURE

FRANCES H. ARNOLD

Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena, CA 91125, U.S.A.

(Received 10 October 1995; accepted 28 January 1996)

**Abstract**—An effective approach to engineering new enzymes is to direct their evolution *in vitro*. By mimicking key processes of Darwinian evolution in the test tube, the functions of enzymes can be explored free from the constraints of function within a living system. Efficient strategies for engineering new enzymes by multiple generations of random mutagenesis and recombination coupled with screening for improved variants have been developed. Our results with industrially important biocatalysts underscore the advantages of this 'evolutionary' approach to protein engineering. (From a talk presented at the first National Academy of Engineering "Frontiers of Engineering" Symposium, 21 September 1995.) Copyright © 1996 Elsevier Science Ltd

### INTRODUCTION

Biological systems are the masters of chemical synthesis. The remarkable specificity of their catalysts, the enzymes, allows hundreds of reactions to proceed simultaneously inside the tiny reactor that is a living cell. Enzymes' ability to carry out complex chemical reactions, and to do so under very mild conditions with virtually no waste products, has earned them the admiration of chemists and biochemists. It is easy to envision that a future chemical industry sensitive to both energy needs and the environment could be modeled after these highly efficient chemical factories.

The molecules responsible for this remarkable performance are the enzymes. Enzymes are proteins, linear chains of typically hundreds of amino acids that fold up into unique, well-defined three-dimensional structures. The backbone of the polymer chain folds into a structure that is unique to the particular catalyst, as illustrated in Fig. 1(a) for the enzyme subtilisin. The enzyme's substrate (gray), the compound on which the reaction is catalysed, fits snugly into the substrate binding pocket. The enzyme positions specific catalytic amino acid side chains (red) where they can assist the chemical reaction to proceed. In Fig. 1(b) the structure of subtilisin showing its amino acid side chains illustrates the complexity of these molecular machines. This complexity allows enzymes to perform the truly impressive functions that support life and create new life. The result of considerable fine-tuning over eons of evolution, this complexity also makes it difficult to manipulate these structures to obtain new and interesting properties.

An enzyme is defined by a unique sequence of amino acids, which in turn is dictated by the organ-

ism's DNA code (the gene) and assembled in the cell (Fig. 2). This amino acid sequence determines how the chain folds and, ultimately, how the enzyme functions. By modifying the amino acid sequence, we can alter the enzyme's function—this field is known as *protein engineering*. Despite intense research into fundamental features governing protein folding and function, there are enormous gaps in our understanding of two critical processes: the relationship between sequence and structure and the relationship between structure and function. As a result, the rational design of new proteins by the classical 'reductionist' approach can be a frustrating exercise indeed. In this article I will introduce a new and highly effective approach to enzyme design and engineering that bypasses the need to understand these processes before embarking on a protein engineering project. But first I will explain why the enzymes provided by nature are not sufficient.

Chemical engineers who try to design real industrial processes using biological catalysts are constantly stymied by a simple fact: biological systems have evolved over billions of years to perform very specific biological functions and to do so within the context of a living organism. Some of the features required for function in a complex chemical network are undesirable when the catalyst is lifted out of context. Conversely, many of the properties we wish an enzyme would have clash with the needs of the organism, or at least were never required. The chemical engineer is hardly impressed by a catalyst whose inability to tolerate the most common of industrial conditions necessitates complicated hardware and reactors of the size of football fields. We need catalysts which are stable to high temperatures, can function in solvents other than water, tolerate wider ranges of pH,

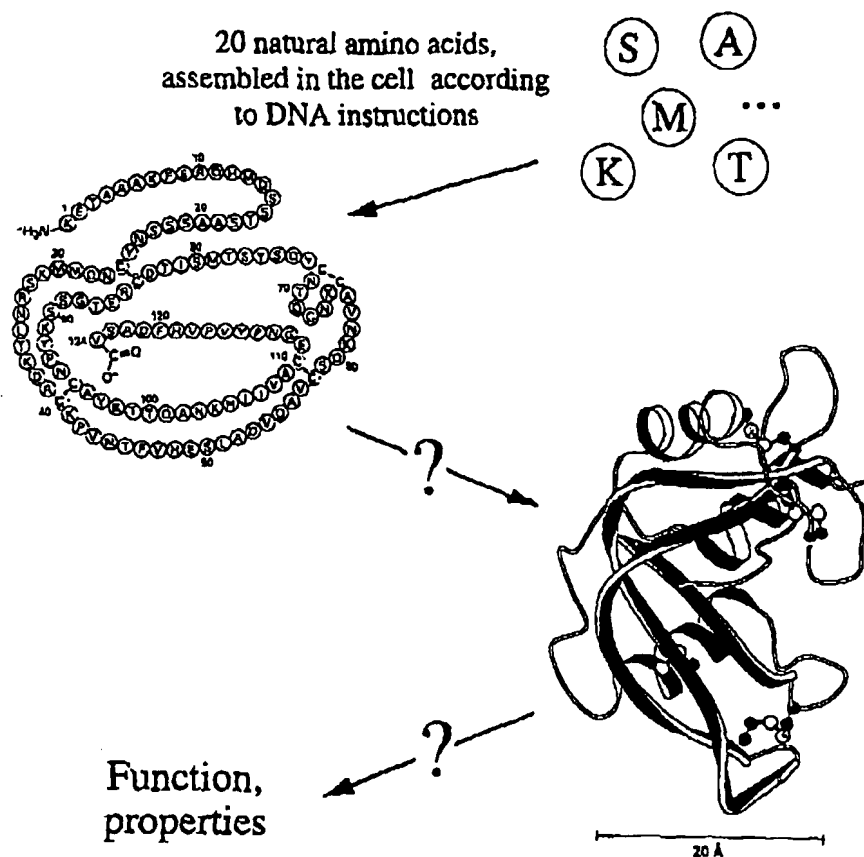


Fig. 2. Protein engineering involves the manipulation of protein structures and functions at the level of the amino acid (or DNA) sequence. Significant gaps in our understanding of the relationships between sequence, structure and function severely limit our ability to 'rationally design' new functions.

catalyse reactions on substrates not encountered in nature, and even catalyse new reactions not found in nature.

Many clues as to how to engineer better enzymes come from studying how nature has created enzymes. By studying the evolution of natural proteins, we have learned in fact that they are highly adaptable, constantly changing molecules, at least over evolutionary time scales. They can adapt to new environments and they can even take on new tasks. We know, for example, that many enzymes catalysing very different reactions have come about by divergent evolution from a common ancestral protein of the same general structure, acquiring diverse capabilities by processes of random mutation, recombination, and natural selection. For example, the versatile protein structure known as the  $\alpha/\beta$  barrel diverged somewhere in the distant past to create a whole series of enzymes we know today (Reardon and Farber, 1995). The four enzymes shown in Fig. 3(a), for example, catalyse quite different reactions; their physical properties and amino acid sequences are also quite disparate. It is useful to note

that, while the barrel-like protein fold is highly conserved, the amino acid sequences and functions of these enzymes are not.

A fascinating recent example of enzyme evolution is the appearance of phosphotriesterase, an  $\alpha/\beta$  barrel enzyme that hydrolyses, at diffusion-limited rates, pesticides and chemical warfare agents that have existed only for about 50 years. It has been suggested that this enzyme, discovered in a soil bacterium, evolved during the last 50 years from a related sequence identified in the common *E. coli* bacterium and now known as the 'phosphotriesterase homology protein' (Scanlan and Reid, 1995). The biological function of this latter protein is unknown.

We also know that enzymes of a given function (for example, all catalysing a particular step in a metabolic pathway) can exhibit widely different properties (stability, solubility, tolerance to pH, etc.), depending on where they are found. For example, the three glyceraldehyde phosphate dehydrogenase (GAPDH) enzymes listed in Fig. 3(b) have very similar three-dimensional structures; their sequences are less similar. We know



Fig. 1. The 275 amino acids of subtilisin E fold into a unique three-dimensional structure. (a) The backbone fold is represented here by a 'ribbon' diagram, constructed from X-ray crystal structure coordinates (Dauter *et al.*, 1991) using the programs MolScript and Raster3D. Peptide substrate and two stabilizing calcium ions are shown in gray. Side chains of catalytic amino acid residues are shown in red. (b) Subtilisin structure showing the positions of the amino acid side chains (yellow).



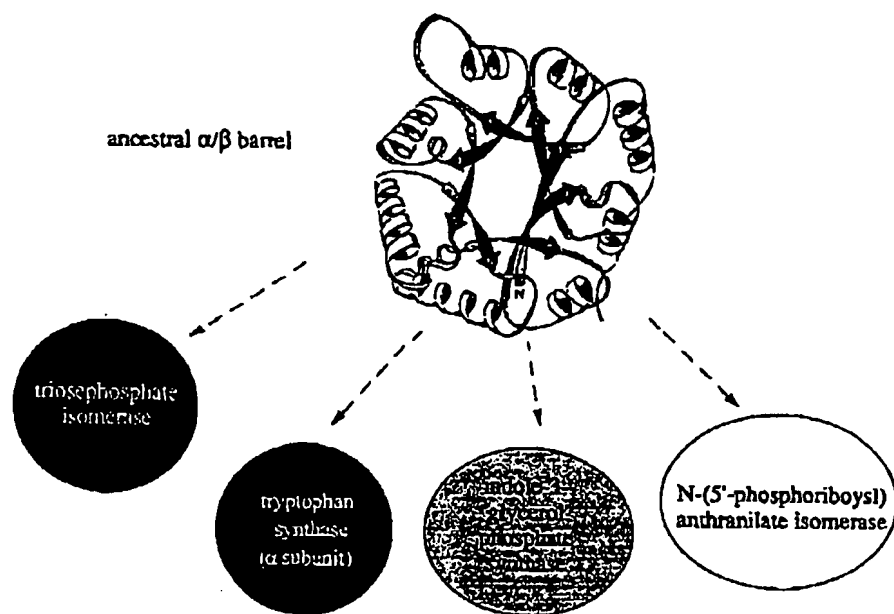
Fig. 6. Molecular model of subtilisin E showing the 12 amino acid substitutions that increase enzyme activity in DMF (You and Arnold, 1996). Yellow amino acids were accumulated during screening for enhanced specific enzyme activity (Chen and Arnold, 1995). Red amino acids were found during screening for total (expressed) enzyme activity (You and Arnold, 1996). Calcium ions and peptide substrate are shown in gray.



Fig. 10. Molecular model of the pNB esterase showing positions of antibiotic *p*-nitrobenzyl ester substrate (yellow), catalytic residues (red), and six beneficial mutations accumulated during directed evolution (orange) (Moore and Arnold, 1996).

that th  
a long  
*Therm*  
temper  
enzym

## a) One enzyme can become another



## b) Enzymes evolve for different environments

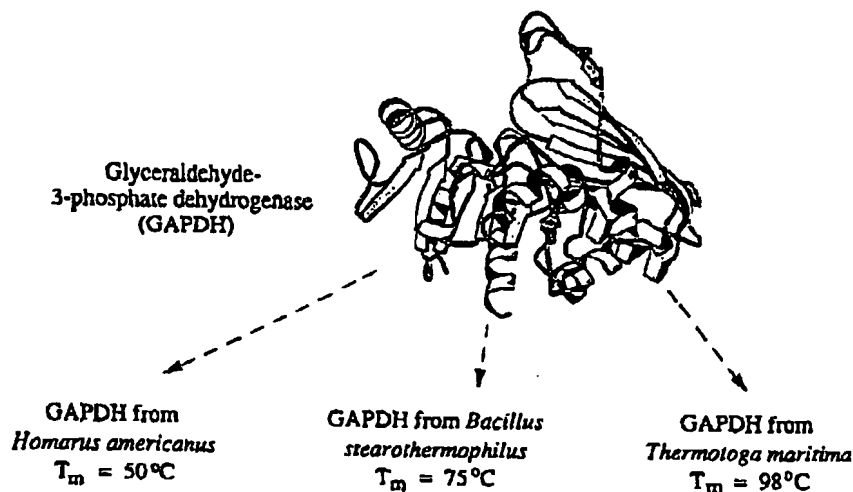


Fig. 3. Structure is conserved during evolution, while amino acid sequences and specific functions are often not. (a) The  $\alpha/\beta$  barrel enzymes indicated appear to have evolved from a common ancestral  $\alpha/\beta$  barrel protein. (b) Three GAPDH enzymes isolated from different organisms have very similar structures, but quite different stabilities and amino acid sequences (Buehner *et al.*, 1974; Skarzynski *et al.*, 1987; Korndorfer *et al.*, 1995).

that they, too, diverged from some common ancestor a long time ago to occupy their current niches. The *Thermotoga maritima* bacterium thrives at very high temperatures in ocean thermal vents; consequently, its enzymes can tolerate much higher temperatures than

the analogous enzymes from an organism which grows under less extreme conditions, such as *B. stearothermophilus*. The *Thermotoga* protein unfolds at  $98^\circ\text{C}$ , while the same enzyme from the American lobster unfolds at only  $50^\circ\text{C}$ . As with the  $\alpha/\beta$  barrel

enzymes, the structural fold of the GAPDHs is highly conserved, while the detailed amino acid sequences and specific properties are not.

#### DIRECTED EVOLUTION: EXPLORING NEW FUTURES

The explosion of tools that has come out of molecular biology during the last 20 years has made it possible for us to consider 'evolving' the components of biological systems—DNA, RNA and proteins—for features never required in nature. We can both speed up the rate and channel the direction of evolution by controlling mutagenesis the rate and types of changes made—and the accompanying 'selection' pressures. As a result, processes that would take millions of years in nature can in principle be accomplished during the time scale of a Ph.D. thesis. By uncoupling the enzymes from the constraints of function within a living system, we can step into and explore a variety of futures, futures that can include novel environments (evolution in a sea of methanol instead of water?) or even entirely new functions (enzymes to break down hazardous chemicals?). We can explore questions such as 'can one catalytic activity become another, and how?' Furthermore, by evolving new functions and thereby new solutions to molecular design problems, we learn things about these amazing molecular machines that might never be revealed if we were to study only those that exist in nature.

The possibilities for biotechnology are especially exciting. Directed evolution is a very practical approach to tailor-making enzymes for a wide range of applications. In addition to building enzymes with new features and functions, we can explore important questions such as 'how might an enzyme change its sequence and properties to break down or evade a drug?' We could conceivably anticipate in laboratory experiments what might happen to drug resistances in nature. In directed evolution experiments we could also tune enzymes to function optimally under conditions specified by us, rather than the context of the living organism in which it evolved. New enzymes could be evolved to carry out reactions never required by living organisms.

#### DEVELOPING A WORKING STRATEGY FOR DIRECTED ENZYME EVOLUTION

In a directed evolution experiment, we first generate a library of many different possible 'solutions' to a problem. The next step is to find the correct solution(s), enzymes that exhibit the desired property. A conceptual challenge comes in planning how to create this library of solutions. The number of possible enzymes one can make is so vast that an exploration of their functions must be carefully guided in order to avoid becoming hopelessly lost. A typical enzyme is a linear polymer of 300 amino acids. With 20 possible amino acids at each position in the chain, there are  $20^{300}$  possible different linear combinations. If even only a small fraction—say, 1 in  $10^{10}$ —of all

these sequences folds into a well-defined three-dimensional structure, there are still more structured proteins than there are atoms in the universe! (Note that even in three billion years, nature has not had a chance to explore but a tiny fraction of the possibilities. This also means that there are very exciting possibilities for future evolution, including evolution in the test tube.) Because a random sampling of amino acid sequences is unlikely to lead to the desired protein, we must begin our exploration by starting from a point that we hope is close to where we want to be—an enzyme that approximates what we want, but is not ideal. Then we evolve it, by accumulating small changes, similar to what happens in nature.

Nature is very good at searching mutant libraries for useful solutions. Unfavorable mutations are winnowed out at the same time as beneficial mutations are amplified, by linking the organism's growth rate and reproductive success to the performance of its components. In this process of *selection*, those organisms which grow faster quickly dominate, allowing an efficient search of very large populations ( $10^9$  or more for bacteria).

Unfortunately, many of the features that are of interest to us cannot be linked to the survival or growth of the host organism—the prerequisite to selection. Enzymes, for example, can tolerate a variety of environments that cannot sustain life, so that the organism dies long before the enzyme has a chance to 'show its stuff'. For most problems of practical interest, in fact, mutant enzyme libraries must be screened rather than selected, one enzyme at a time. That is, the enzyme variants must be tested individually (screened) for the property of interest. This unfortunate reality effectively limits the search for improvements to mutant libraries containing perhaps  $10^4$ – $10^6$  variants, several orders of magnitude smaller than what one can search when survival depends on success.

The strategy for molecular evolution is then illustrated by calculating how many different sequences one can create by starting from a given enzyme and making a few amino acid substitutions, as shown in Table 1. While there are only 5700 possible single mutants of a 300 amino acid enzyme, there are still more than 30 billion different sequences that differ from the original enzyme at only three positions. While a rapid screen might be able to cover a large fraction of all single mutants, and even some significant fraction of all double mutants, screening would be unable to give more than a very sparse sampling of the enzymes with multiple amino acid substitutions. Unless a vast majority of the mutations led to the desired property, dealing with a library of multiple mutations would be an experiment based on wishful thinking! (As might be expected for a finely tuned molecular machine, most mutations are deleterious or at least neutral; beneficial mutations are generally rare. The frequency with which one can expect to find beneficial mutations will depend on the extent to

which they have been optimally approached basis for effectively; therefore addition, mutation amino acid

The problem is to be the evolutionary organic is normally the peptide will also bond for participate hydrolysis; folded and polar or (DMF). low. The subtilisin unhappy number ten—prebalance the dissolved complex not approach. and asked would find 1993; You

The approach for directed a new function Fig. 4. I a function in aqueous indeed. S in DMF.

Table

N

Note: S number of is (19)M



which the particular feature of interest has already been optimized: the pathways up the mountain necessarily decrease in number as the pinnacle is approached.) Because luck is generally not an acceptable basis for the success of an experiment, the search is effectively limited to proteins with sequences and therefore properties very similar to their parents. In addition, we must be able to tune the rate or mode of mutation to produce libraries of primarily single amino acid substitutions.

The principles and power of directed evolution are best illustrated with examples. The first example will be the evolution of an enzyme to function in a polar organic solvent. It is well known that subtilisin, which normally cuts up peptides and proteins by cleaving the peptide bonds linking the amino acids together, will also catalyse peptide bond formation. Peptide bond formation is favored in organic media, as water participates in unwanted side reactions as well as hydrolysis of the product. Subtilisin actually remains folded and reasonably stable in high concentrations of polar organic solvents such as dimethylformamide (DMF). Unfortunately, the catalytic activity is very low. There is no fundamental reason, however, why subtilisin could not function in DMF—the enzyme's unhappiness reflects a balance among a very large number of noncovalent interactions in the system—protein, solvent, substrates and products—a balance that is adversely affected when the protein is dissolved in a nonaqueous medium. Because these complex interactions are poorly understood, we could not address this problem by a rational design approach. We therefore took the 'irrational' approach and asked whether we could 'evolve' a subtilisin that would function well in DMF (Chen and Arnold, 1991, 1993; You and Arnold, 1996).

The arguments set out above led us to the strategy for directing the evolution of an enzyme to perform a new function (or, in this case of subtilisin, an old function but under new conditions) illustrated in Fig. 4. In comparison to the enzyme performing a function for which it is selected, peptide hydrolysis in aqueous media, the new job is performed poorly indeed. Subtilisin has not been selected for hydrolysis in DMF, and there is, not surprisingly, a great deal of

room for improvement. Because it is feasible to search only those subtilisin mutants with one or two amino acid substitutions, we will create and screen a library of such mutants for progeny slightly better than their parent. The screening method for identifying useful mutations should ensure that the expected small enhancements brought about mainly by single mutations can be measured. Although these progenies will generally resemble their parents, after many generations new features can develop, such that the descendants can be quite different from their ancestor. Therefore, the generation of new, useful enzymes also relies on having an effective strategy for accumulating many such small improvements. One such strategy involves carrying out sequential generations of random mutagenesis on the gene (DNA sequence coding for the enzyme) to create a mutant library, coupled with screening of the resulting proteins. In each generation a single variant is chosen as the parent for the next generation, and sequential cycles allow the evolution of the desired features.

We implemented this strategy to evolve subtilisin to function in DMF. A powerful molecular biology tool, the polymerase chain reaction (PCR), was used to make millions of copies of the gene that codes for the natural, or wild-type enzyme. By carrying out this (enzymatic) reaction under sub-optimal conditions, we could introduce base substitutions randomly throughout the DNA at a controllable rate. At the end of this reaction we have millions of gene copies, most slightly different from the wild-type one. These genes are placed back into a circular double-stranded piece of DNA (a plasmid) that contains all the instructions the bacterial cells need to translate the DNA into protein. When the bacteria are transformed with these plasmids, we have millions of individual chemical factories, each producing a different variant of the original enzyme.

Table 1. The molecular evolution 'number problem'

No. of amino acid changes	No. of possible variants
0	1
1	5700
2	16,190,850
3	30,557,530,900
4	43,109,036,717,100
5	48,489,044,499,400,000

Note: Starting with an enzyme of 300 amino acids, the number of sequences containing  $M$  amino acid substitutions is  $(19)M \cdot 300! / [(300 - M)!M!]$ .

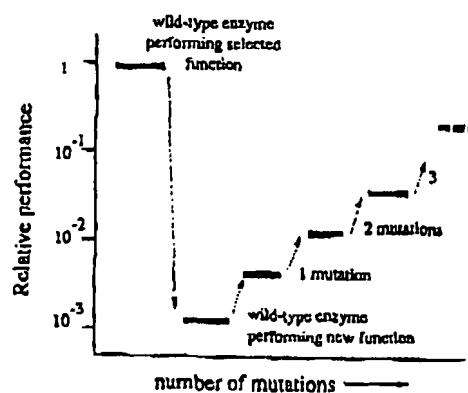


Fig. 4. A working strategy for directed enzyme evolution. The screening method should ensure that small enhancements brought about mainly by single mutations can be measured. The evolution of a new, useful enzyme requires an effective strategy for accumulating many such small improvements.

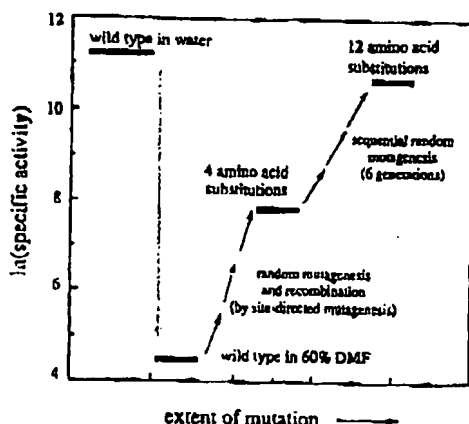


Fig 5. Results of directed evolution of subtilisin for activity in DMF by sequential generations of random mutagenesis and screening. The accumulation of 12 amino acid substitutions in sequential generations of random mutagenesis and screening resulted in an enzyme >500-fold more active than the wild-type enzyme in 60% DMF.

Next, the bacterial colony or colonies which produce a subtilisin that is more active in DMF must be found. In this early experiment our screening strategy was crude, but effective. Because subtilisin is secreted from the bacilli, the variants could be screened visually on nutrient plates containing a protein (casein), in the presence and absence of DMF. The active enzyme creates a visible 'halo' surrounding the bacterial colony whose size is proportional to the catalytic activity.\* Variants with higher hydrolysis activity than wild-type on the DMF-containing plates could be identified from their bigger halos (Chen and Arnold, 1991).

The results of the directed evolution effort are summarized in Fig. 5. At first we identified three amino acid substitutions that individually improved the wild-type enzyme's activity several-fold. Using site-directed mutagenesis we combined those three with a fourth mutation reported to improve activity and stability in other subtilisins, to obtain a four amino acid variant about 40-fold more active than wild-type in 60% DMF (Chen and Arnold, 1991). Since the process of sequencing the genes of all the positive variants and then combining the mutations by site-directed mutagenesis was laborious, we decided to carry out sequential generations of random mutagenesis and screening, no longer stopping on the way to sequence the intermediates. Applying an additional six generations of mutagenesis and screening a few hundred colonies in each generation, we created an

\*Because halo size also depends on enzyme expression level, enzyme diffusion and colony size, it is useful for a 'rough cut'. Positives were confirmed by a second level of screening in liquid culture (Chen and Arnold, 1993).

enzyme that is more than 500-fold more active in 60% DMF than the wild-type subtilisin E (You and Arnold, 1996). This enzyme exhibits substantial activity even in 85% DMF. The whole process was surprisingly rapid: a total of only about 10,000 colonies were screened to obtain a huge improvement in catalytic activity.

The gene for the final evolved enzyme was sequenced to determine the amino acid substitutions that allowed this enzyme to recover its activity in DMF. Of 275 amino acids, 12 were altered; their positions are indicated in Fig. 6. Although the DNA substitutions are targeted randomly throughout the entire subtilisin gene sequence, the amino acid substitutions that enhance catalytic activity are all positioned on the surface of the enzyme, surrounding the active site and substrate binding pocket. The majority are in evolutionarily variable loops that connect elements of conserved secondary structures (helices and sheets) (Chen and Arnold, 1993; You and Arnold, 1996). This information could of course be utilized in developing more 'rational' design strategies, including narrowing the sequences exposed to random mutagenesis in directed evolution.

Finally, it is worth noting that the resulting enzyme is indeed a far more efficient catalyst than wild-type subtilisin for the polymerization of amino acids. This evolved enzyme can catalyze, for example, the formation of poly-L-methionine starting from a racemic mixture of methionine methyl ester. The evolved enzyme allows the synthesis of significantly longer polymers and at much higher yields than the native enzyme in 60–70% DMF (Zhao, H. unpublished results).

The advantage of directed evolution over site-directed mutagenesis is clear: the same amount of effort could support the construction and screening of at most a few dozen variants with mutations directed to specific locations. Without a clear mechanism, it would be difficult indeed to pinpoint 12 amino acid substitutions that enhance catalytic activity in DMF. Even then, single site-directed mutations would have to be accumulated to create a useful enzyme, itself a substantial mutagenesis effort involving trial and error to find optimal combinations.

The most attractive feature of the evolutionary strategy outlined in Fig. 4 is its simplicity. It is possible, however, that this simple 'up-hill climb' approach is not an optimal approach to the evolution of a particular enzyme. There are obviously a great number of pathways possible for the evolution of a protein, and each choice of parent for the next generation represents an irreversible step along one particular pathway. What would happen if we simply repeated the experiment? Depending on which pathway was chosen or which mutation happened to be found first, the enzyme could end up on a local optimum, unable to evolve further. This approach may also appear slow: improvements are small in each step and necessarily become harder to find the closer the enzyme gets to an optimum.

An ai recently vantage Gene re genes on speed of benefici remove tion into genes w combin directed We h ating an sis of tl antibiot ing grou of cep moval ) recovery large ar a major effort so perform al., 1971 activity by screc the enz; requirec competi We v esterase presence achieve to belie natural unknow pNB es very scr tures w setting, through The v cells in ' tion tha screenin used fo hydrolysis formant able for therefor ilar, but order to screenir wells of spectroi ance in Using colonies

## SEX IN THE TEST TUBE

An alternative directed evolution strategy we have recently explored incorporates some important advantages attributed to sex in the evolutionary process. Gene recombination, the cutting and pasting of whole genes or pieces of genes, can significantly increase the speed of molecular evolution by rapidly accumulating beneficial mutations and providing a mechanism to remove deleterious ones. To incorporate recombination into directed evolution, we randomly recombine genes with positive mutations. A search for better combinations of mutations completes a generation of directed evolution.

We have tested this new 'sexual' approach by creating an enzyme that efficiently catalyses the hydrolysis of the *p*-nitrobenzyl (PNB) ester of a  $\beta$ -lactam antibiotic in the presence of DMF. The PNB protecting group is often used during the large-scale synthesis of cephalosporin-type antibiotics. Its selective removal presents problems, however, particularly for recovery and disposal of the zinc catalyst and the large amounts of organic solvents used. Therefore, a major pharmaceutical company devoted significant effort some years ago to finding an enzyme that would perform this selective hydrolysis reaction (Brannon *et al.*, 1976; Zock *et al.*, 1994). An enzyme with some activity towards PNB ester hydrolysis was identified by screening a large number of microorganisms, but the enzyme's low activity, especially in the solvents required to solubilize these materials, made it a poor competitor to the classical chemical catalyst.

We were challenged in 1994 to evolve a PNB esterase with much higher activity, particularly in the presence of the polar organic solvents required to achieve high substrate solubility. We had two reasons to believe that this could be done. First, the enzyme's natural function and, therefore, natural substrates are unknown, but they are unlikely to be the antibiotic PNB esters. Second, the natural enzyme's activity is very sensitive to organic solvents. Because these features were never required in the enzyme's natural setting, we could expect considerable improvement through directed evolution.

The wild-type esterase is not secreted by the *E. coli* cells in which it is made, nor does it carry out a reaction that is easily measured. Thus, we had to develop screening strategies more sophisticated than those used for the subtilisin. The *p*-nitrobenzyl ester hydrolysis reaction is assayed laboriously by high performance liquid chromatography, a method unsuitable for screening tens of thousands of colonies. We therefore devised a rapid screening assay using a similar, but not identical, *p*-nitrophenyl ester substrate, in order to have an easy-to-read colorimetric signal. The screening reactions could then be carried out in the 96 wells of a plastic microtiter plate, using an automatic spectrophotometer to read and analyse the absorbance in all 96 wells at once.

Using this rapid assay to screen about a thousand colonies per generation, we completed several sequen-

tial cycles of random PCR mutagenesis and screening, as illustrated in Fig. 7 (Moore and Arnold, 1996). After four generations, the enzyme's specific activity in 15% DMF had improved 15-fold. In the fourth generation, we collected not one, but 64 different clones, some of which were better than the parent, and many of which were not. The purpose for this was two-fold. First we wanted to make sure that our screening strategy was working properly to give us an enzyme that would catalyse the desired *p*-nitrobenzyl hydrolysis reaction, not only the colorimetric *p*-nitrophenyl screening reaction. The activities of each of the 64 clones in both reactions are compared in Fig. 8.

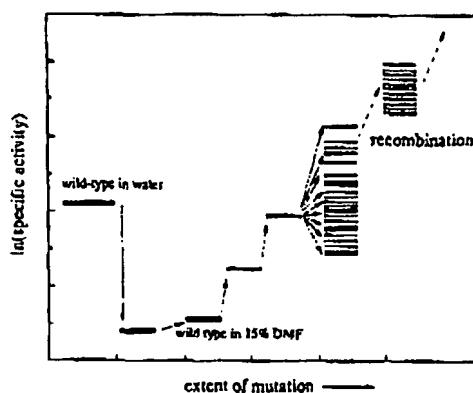


Fig. 7. Directed evolution of pNB esterase in 15% DMF involved four generations of random mutagenesis and screening, followed by one round of recombination of the five best genes from generation 4. The best variant obtained after four generations is 15-fold more active than wild-type. The best variant from screening 400 colonies of the recombination pool is ~30-fold more active than wild-type.

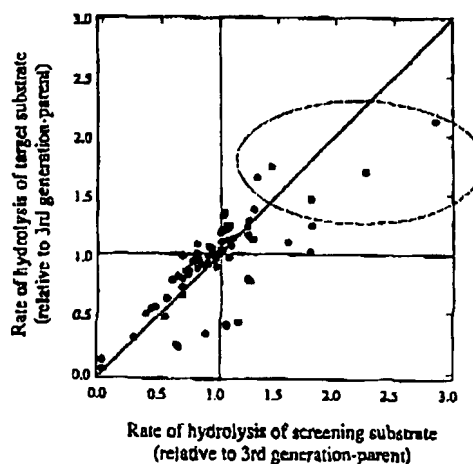


Fig. 8. Comparison of activities on target (*p*-nitrobenzyl) and screening (*p*-nitrophenyl) substrate of 64 pNB esterase variants isolated after fourth generation of random mutagenesis and screening, relative to parent enzyme from the third generation. The five most active variants (inside oval) were pooled for random recombination (see Fig. 9).

If the screening reaction perfectly mimicked the desired reaction, all the points would lie on the 45° line. Although somewhat scattered, there is nonetheless a reasonable correlation: the rapid screen provides an indication of evolution of the desired activity that is acceptable for making a rough cut of positive clones.

The second reason for studying this group of variants was to test the alternate, sexual approach for accumulating effective mutations. We thus collected the five best mutants, those in the dotted oval in Fig. 8, and recombined them using a 'sexual' PCR method recently described by Stemmer (1994a,b). How the genes are randomly recombined is shown schematically in Fig. 9(a). The genes are pooled in the test tube and fragmented with an enzyme that cuts the DNA at random positions. In Fig. 9(b), the polyacrylamide electrophoresis gel that separates the DNA fragments by length shows that the DNA has been digested into a smear of different-sized pieces. We collected the fragments 200–300 base pairs in length by extracting the DNA from the appropriate piece of gel. The full-length gene can be reassembled from this pool of random fragments, again using the PCR technology, to create a new gene library in which the mutations were present in their different possible combinations. These reassembled, recombined genes were inserted back into the plasmid and expressed in the *E. coli*. The best of those recombined genes were identified, as before, by screening the enzymes they code for and produce in the microorganism.

Screening only ~400 colonies yielded eight clones with activity significantly greater than the best of the five parents—this yield of positives is at least 20-fold higher than we found by screening the genes with point mutations alone (typically 1/1500). Recombination can enhance directed evolution by making use of the information present in a population of improved enzymes produced by mutagenesis and screening, information that would otherwise be discarded. Thus far, we have improved the enzyme's specific activity towards the antibiotic substrate more than 30-fold in 15% DMF. The total expressed activity is at least 50-fold greater than the original system we started with.

Sequencing of the genes coding for improved enzymes once again allowed us to identify the amino acid substitutions responsible for the observed improvements in catalytic performance. Six effective mutations are illustrated in Fig. 10, on a model of the pNB esterase developed from the X-ray crystal structure of a homologous enzyme (Moore and Arnold, 1996). As for the case of subtilisin, most of the mutations are at or near the solvent-accessible surface. Only one of the six is deeply buried. In contrast to subtilisin, however, none of the effective amino acid substitutions lie in segments of the esterase predicted to interact directly with the bound substrate. It is possible that the homology modeling yielded an incorrect structure, and the mutations do interact with the *p*-nitrobenzyl substrate. Or, it may be that the amino acid substitutions sampled at positions adjacent

to the substrate were all deleterious, and small improvements were only obtained by altering amino acids further away. In any case, the mechanism(s) by which these amino acid substitutions enhance the catalytic activity of the evolved pNB esterases are subtle and would have been very difficult to predict in advance.

## CONCLUSIONS

The directed evolution approach clearly allows us to engineer enzymes with novel functions and features. In contrast to 'rational' design approaches, directed evolution can be applied even when very little is known about an enzyme's structure or catalytic mechanism. Since the vast majority of proteins remain largely uncharacterized, this marks a huge advantage for the evolutionary methods. This approach, because it allows us to explore novel solutions to protein design problems, also promises to teach us a great deal about protein structure and function.

Future research in directed evolution will include development of large-scale screening methods, so that efficient searches of large mutant libraries can be performed. The construction of optimized mutant libraries will also decrease the need for screening. In addition to streamlining efforts to 'tune' enzymes, these improvements will allow larger leaps—such as the evolution of new catalytic activities—to take place. Significant improvements in the ease and power of directed evolution will also come from optimizing the search strategies. The many similarities to optimization problems in other fields make this a fertile ground for collaborative efforts among theoreticians and experimentalists from a wide range of engineering disciplines.

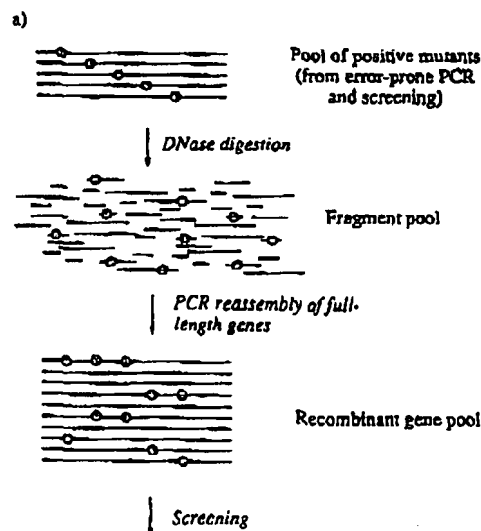


Fig. 9. Recombination of mutations by gene shuffling. (a) 'Sexual' PCR method (Stemmer, 1994a,b) involves random digestion of the gene pool using DNase enzyme, followed by gene reassembly using PCR. Reassembled genes contain the different combinations of mutations.

Fig. 9. (1  
200–300

Acknowl  
postdoc  
especiall  
This wo  
search a  
Biologic  
Office o  
Renewal

Buehner  
Rossn  
n-glyc  
Biol. S  
Brannor  
esterif  
micro  
Chen, F  
of ar.

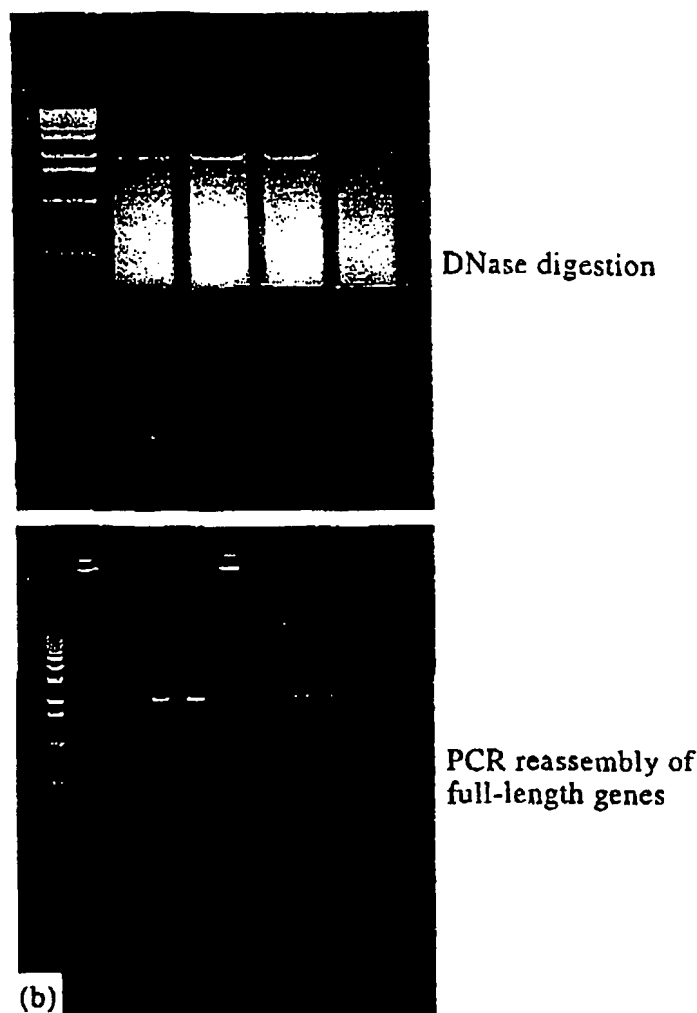


Fig. 9. (b) Top polyacrylamide electrophoresis gel shows the separation of the digested gene fragments by size. Fragments 200–300 base pairs long were recovered by extracting the excised gel segment. These were reassembled into the full-length gene (bottom gel, lanes 3–5 and 6–8). First lane on left is a 'ladder' of DNA of known molecular weights.

**Acknowledgements**—I am grateful to all the students and postdoctoral researchers who have contributed to this effort, especially Jeffrey Moore, Kevin Chen and Huimin Zhao. This work was supported by the U.S. Office of Naval Research and the U.S. Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables.

#### REFERENCES

- Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. and Rossmann, M. G., 1974. Three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.* **90**, 23–49.
- Brannon, D. R., Mahe, J. A. and Fukuda, D. S., 1976. Deacetylation of cephalosporin para-nitrobenzyl esters by microbial enzymes. *J. Antibiotics* **29**, 121–124.
- Chan, K. and Arnold, F. H., 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5618–5622.
- Chen, K. and Arnold, F. H., 1991. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnology* **9**, 1073–1077.
- Dauter, Z., Betzel, C., Genov, N., Pipon, N. and Wilson, K. S., 1991. Complex between the subtilisin from a mesophilic bacterium and the leech inhibitor eglin-C. *Acta Crystallogr.* **B47**, 707–730.
- Komdoerfer, L., Sticpe, B., Huber, R., Tomsch, A. and Jaenicke, R., 1995. The crystal structure of hologlycerinaldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511–521.
- Moore, J. and Arnold, F. H., 1996. Directed evolution of a *p*-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnol.* **14**, 458–467.
- Reardon, D. and Farber, G. K., 1995. Protein motifs 4. The structure and evolution of  $\alpha/\beta$  barrel proteins. *FASEB J.* **9**, 497–503.



- Scanlan, T. S. and Reid, R. C., 1995, Evolution in action. *Chem. Biol.* 2, 71-75.
- Skarzynski, T., Moody, P. C. E. and Wonacott, A. J., 1987, Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* 193, 171-187.
- Stemmer, W. P. C., 1994a, DNA shuffling by random mutagenesis and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10,747-10,751.
- Stemmer, W. P. C., 1994b, Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 340, 389-391.
- You, L. and Arnold, F. H., 1996, Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Engng* 9, 77-83.
- Zock, J., Cantwell, C., Swartling, J., Hodges, R., Pohl, T., Sutron, K., Rostek, P. Jr., McGilvray, D. and Queener, S., 1994, The *Bacillus subtilis* *pnbA* gene encoding *p*-nitrobenzyl esterase: cloning, sequence and high-level expression in *E. coli*. *Gene* 151, 37-43.

The rec:  
(c.g. vac  
a cell-di  
the host  
ccss-sca  
(Middel  
mechan  
used me  
and Ku  
Num  
as majc  
homoge  
Marcus  
pingem  
1981), r  
(pressur  
shear (A  
elling a  
they ha  
homoge  
simplifi  
informa  
a homo  
and flo  
previou  
The r  
pressur  
Gaulin  
ted witl  
remove  
is used

## Site-directed mutagenesis of residues 164, 170, 171, 179, 220, 237 and 242 in PER-1 $\beta$ -lactamase hydrolysing expanded-spectrum cephalosporins

Anne-Typhaine Bouthors<sup>1</sup>, Jean Delettre<sup>2</sup>,  
Pierre Mugnier<sup>1,3</sup>, Vincent Jarlier<sup>1</sup> and  
Wladimir Sougakoff<sup>1,4</sup>

<sup>1</sup>Laboratoire de Recherche Moléculaire sur les Antibiotiques, Université Pierre et Marie Curie (Paris VI), Faculté de Médecine Pitié-Salpêtrière and Faculté de Médecine, Broussais-Hôtel Dieu, F-75634 Paris cedex 13 and,

<sup>2</sup>Laboratoire de Mineralogie-Cristallographie, CNRS URA 09, Université Pierre et Marie Curie (Paris VI), F-75252 Paris cedex 05, France

<sup>3</sup>Present address: Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK

<sup>4</sup>To whom correspondence should be addressed. E-mail: sougakoff@lmcp.jussieu.fr

The class A  $\beta$ -lactamase PER-1, which displays 26% identity with the TEM-type extended-spectrum  $\beta$ -lactamases (ESBLs), is characterized by a substrate profile similar to that conferred by these latter enzymes. The role of residues Ala164, His170, Ala171, Asn179, Arg220, Thr237 and Lys242, found in PER-1, was assessed by site-directed mutagenesis. Replacement of Ala164 by Arg yielded an enzyme with no detectable  $\beta$ -lactamase activity. Two other mutants, N179D and A164R+N179D, were also inactive. Conversely, a mutant with the A171E substitution displayed a substrate profile very similar to that of the wild-type enzyme. Moreover, the replacement of Ala171 by Glu in the A164R enzyme yielded a double mutant which was active, suggesting that Glu171 could compensate for the deleterious effect of Arg164 in the A164R+A171E enzyme. A specific increase in  $k_{cat}$  for cefotaxime was observed with H170N, whereas R220L and T237A displayed a specific decrease in activity towards the same drug and a general increase in affinity towards cephalosporins. Finally, the K242E mutant displayed a kinetic behaviour very similar to that of PER-1. Based on three-dimensional models generated by homology modelling and molecular dynamics, these results suggest novel structure–activity relationships in PER-1, when compared with those previously described for the TEM-type ESBLs. **Keywords:**  $\beta$ -lactamase/expanded-spectrum cephalosporins/homology modelling/PER-1/serine enzyme

### Introduction

PER-1 is a class A  $\beta$ -lactamase characterized by a high catalytic activity against expanded-spectrum cephalosporins [e.g. cefotaxime (CTX) and ceftazidime (CAZ)] and monobactams [e.g. aztreonam (AZT)]. The enzyme, which was first identified in *Pseudomonas aeruginosa* (Nordmann *et al.*, 1993), displays a kinetic behaviour very similar to that of various extended-spectrum  $\beta$ -lactamases (ESBLs) belonging to the TEM and SHV families (Jacoby and Medeiros, 1991). In contrast, PER-1 shares a relatively low amino acid identity with these latter enzymes, e.g. 26% with TEM-3 (Nordmann and Naas, 1994).

Recently, we have undertaken biochemical studies in order

to elucidate the molecular basis of PER-1 activity against expanded-spectrum cephalosporins (Bouthors *et al.*, 1998). Molecular modelling and site-directed mutagenesis were used to investigate in this enzyme the role played by the amino acid residues corresponding to those found at positions 104, 164, 238 and 240 in the TEM-type ESBLs. In brief, two residues, Asn104 and Ala164, were shown to be important for the activity of PER-1. Asn104, which corresponds to the lysine residue found at the same position in various TEM-type ESBLs, would be connected to the key catalytic residue Glu166 via a hydrogen bond network, whereas Ala164, which corresponds to a highly conserved arginine in the  $\Omega$ -loop of class A  $\beta$ -lactamases described so far (Ambler *et al.*, 1991), could play an important structural role. By contrast, modification of the serine residue found at position 238 in PER-1, which is an amino acid found specifically in a large number of TEM-type ESBLs (Bush and Jacoby, 1997), resulted in no significant modification of the activity of PER-1 against expanded-spectrum cephalosporins. Similarly, Gly240 in PER-1 was shown to have no essential role in the substrate profile of the enzyme. Finally, the catalytic residue Glu166, found in all class A  $\beta$ -lactamases, appeared to be essential to the  $\beta$ -lactamase activity of PER-1. However, an unexpected residual activity against CAZ and AZT was observed for a mutant in which Glu166 was replaced by Ala, suggesting that other residues in PER-1 could contribute to the high activity of the enzyme against expanded-spectrum cephalosporins.

In this work, we investigated other amino acid residues found either within or at the vicinity of the PER-1 active site: Ala164, His170, Ala171 and Asn179 which are located within the putative  $\Omega$ -loop of PER-1, Thr237 which is found at the end of the  $\beta$ 3 strand and is likely to participate in the formation of the oxyanion hole (Herzberg and Moulton, 1987; Strynadka *et al.*, 1992), Arg220 which is located in a position similar to that of Arg244 found on strand  $\beta$ 4 in TEM-1 and which could contribute to the stabilization of the oxyanion pocket via hydrogen bonding interactions with strand  $\beta$ 3 (Moews *et al.*, 1990; Jacob-Dubuisson *et al.*, 1991) and Lys242 found in the loop connecting strands  $\beta$ 3 and  $\beta$ 4. All these residues were modified by site-directed mutagenesis and the kinetic properties of the resulting mutants were characterized. By using homology modelling and molecular dynamics simulations, we have attempted to interpret at the structural level the kinetic data obtained for some of the  $\Omega$ -loop mutants.

### Materials and methods

#### Chemicals

Antibiotic powders were provided by the following manufacturers: penicillin G, Laboratoires de Thérapeutique Moderne (Suresnes, France); ampicillin, cephalothin and kanamycin, Sigma Chemical (St Louis, MO, USA); cefotaxime, Laboratoires Roussel (Paris, France); nitrocefin and ceftazidime, Glaxo (Paris, France); and aztreonam, Bristol Myers Squibb (Paris-La Defense, France).

**Table 1.** Nucleotide sequence of the oligonucleotides used in site-directed mutagenesis

Amino acid modification	Oligonucleotide sequence <sup>a</sup>
Ala164 → Arg	5'-CATCTGCGCTTCATTT <u>CGG</u> ACCACAGCGGTCTC-3'
His170 → Asn	5'-CACCTGATCATCGGCGTTCATCTGCGCTTCATT-3'
Ala171 → Glu	5'-CTGCACCTGATCATCTT <u>CGT</u> GCGATCTGCGCTTC-3'
Asn179 → Asp	5'-TTTCATCGAGGTCCAGTCTTGATACTGCACCTG-3'
Arg220 → Leu	5'-TAACAAACCTTTTAACAGCTCTGGTCCTGTGGT-3'
Thr237 → Ala	5'-GGCTTTGATACCCGAAGCACCAGTTTATGTG-3'
Lys242 → Glu	5'-CGCAGTTTTCGCGCTTGATACCCGAAGTACC-3'

<sup>a</sup>Specific base changes are underlined.

The restriction enzymes used in this study were obtained from Boehringer Mannheim (Meylan, France) and T4 DNA ligase from Promega (Madison, WI, USA). [<sup>32</sup>P]dCTP was purchased from Isotopchim (Ganagobie, France).

#### *Escherichia coli strains, plasmids and growth conditions*

*E. coli* CJ236 (Kunkel *et al.*, 1987) and MV1190 (McClary *et al.*, 1989) were used as hosts for phages in site-directed mutagenesis experiments. *E. coli* JM109 (Promega) was used for DNA cloning experiments and for expression of *bla*<sub>PER-1</sub> and the corresponding mutant genes.

The recombinant plasmid pRAZ1, encoding *bla*<sub>PER-1</sub>, has been described by Nordmann *et al.* (1993). Bacteriophage M13mp19 (Messing, 1983) was used as a vector in site-directed mutagenesis experiments. Plasmid pK19 (kanamycin<sup>R</sup>) (Pridmore, 1987) was used in cloning experiments.

*E. coli* MV1190 and JM109 were grown at 37°C in Luria-Bertani (LB) (Difco, Detroit, MI, USA) and brain-heart infusion (BHI) (Difco), respectively. Solid media were obtained by the addition of 2% Bacto-Agar (Difco). Kanamycin (25 µg/ml) and ampicillin (100 µg/ml) were added when necessary. Competent *E. coli* cells were prepared and transformed as described by Chung *et al.* (1989).

#### *Nucleic acid techniques*

Plasmid DNA was purified using either the alkaline lysis for mini-preparations (Bimboim and Doly, 1979) or the Qiagen plasmid kit for maxi-preparations (Qiagen, Hilden, Germany). Isolation of single-stranded DNA and other standard DNA manipulations were carried out according to Sambrook *et al.* (1989). Double- and single-stranded DNA sequencing were carried out by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using the T7 Sequencing kit (Pharmacia Biotech, Saint Quentin en Yvelines, France).

#### *Site-directed mutagenesis*

Site-directed mutagenesis experiments were performed as described previously (Bouthors *et al.*, 1998). In brief, the *bla*<sub>PER-1</sub> gene was excised from pRAZ1 (1.3 kb) and introduced into M13mp19 RF. Site-directed mutagenesis was performed using the uracil template procedure of Kunkel *et al.* (1987). The sequences of the synthetic phosphorylated oligonucleotides (Eurogentec, Liege, Belgium) used to introduce the different mutations in the *bla*<sub>PER-1</sub> gene are listed in Table 1. After mutagenesis, each mutant gene was cloned into plasmid pK19 and the recombinant plasmids thus obtained were introduced by transformation into *E. coli* JM109. The mutant genes were all sequenced in their entirety and on both strands.

#### *Expression and purification of the wild-type and mutant β-lactamases*

The wild-type and mutant enzymes were purified from 1 l cultures by a two-step procedure based on an anion-exchange column followed by a gel filtration, as described previously (Bouthors *et al.*, 1998). The mutant β-lactamases displaying significant activity were detected using the chromogenic cephalosporin nitrocefin (O'Callaghan *et al.*, 1972), while the almost inactive enzymes A164R, N179D and A164R+N179D were identified by electrophoresis on 12% SDS-polyacrylamide gels (Laemmli, 1970), with the wild-type PER-1 β-lactamase as a molecular mass reference. In order to avoid concerns about enzyme stability, kinetic studies were performed shortly after purification. The purity of the different enzymes was assessed by Coomassie Blue staining of SDS-polyacrylamide gels after electrophoresis. Protein concentration was determined by measuring the absorbance at 280 nm (Lorber and Giege, 1992) with an  $\epsilon$  value of 34 850 M<sup>-1</sup>.cm<sup>-1</sup> (Bouthors *et al.*, 1998). For mutants exhibiting more than one protein band on SDS-PAGE analysis, the intensity of the β-lactamase band was measured with a computerized densitometer (Densylab, Bioprobe) and the enzyme concentration was determined with reference to a standard BSA scale analyzed in the same conditions.

#### *Isoelectric focusing*

Isoelectric focusing was performed with a LKB Multiphor apparatus with pH 3.5–9.5 PAG plates (Pharmacia Biotech). Gels were focused at 30 W for 90 min at 10°C. β-Lactamase activity was revealed by staining with the nitrocefin assay.

#### *Determination of the kinetic parameters of the wild-type and mutant enzymes*

Kinetic assays were performed spectrophotometrically in 0.1 M sodium phosphate buffer (pH 7.0) at 30°C on a Uvikon 940 spectrophotometer. The wavelengths and the extinction coefficients used were as follows: penicillin G, 232 nm,  $\Delta\epsilon = -1100$  M<sup>-1</sup>.cm<sup>-1</sup>; cephalothin, 262 nm,  $\Delta\epsilon = -7960$  M<sup>-1</sup>.cm<sup>-1</sup>; cefotaxime, 260 nm,  $\Delta\epsilon = -6710$  M<sup>-1</sup>.cm<sup>-1</sup>; ceftazidime, 260 nm,  $\Delta\epsilon = -8660$  M<sup>-1</sup>.cm<sup>-1</sup>; and aztreonam, 318 nm,  $\Delta\epsilon = -650$  M<sup>-1</sup>.cm<sup>-1</sup>. For each antibiotic, initial rates were measured at six different substrate concentrations. Kinetic parameters were determined by fitting the Michaelis-Menten equation to the experimental data using the regression analysis program LEONORA written by Cornish-Bowden (1995). The values for  $k_{cat}$  and  $K_m$  were estimated using a non-linear least-squares regression method with dynamic weights (Cornish-Bowden, 1995).

#### *Molecular modelling*

The refined theoretical three-dimensional structures of PER-1 and the mutant enzymes were constructed by homology modelling using the computer program Swiss-Model (Peitsch, 1996), as described previously (Bouthors *et al.*, 1998). The models were then subjected to 5000 steps of energy minimization using the Powell minimizer of X-PLOR (Brunger, 1988). The Ω-loop region in the resulting minimized structures was subjected to molecular dynamic simulations in vacuum. The molecular dynamics were initially performed on the 150–190 region of PER-1 containing the Ω-loop (residues 161–179) and the two α-helix regions enclosing the loop (residues 150–160 and 180–190, respectively). The results obtained from this large segment indicated that the two α-helix regions enclosing the loop were very stable (r.m.s.d. = 0.2 Å).



Therefore, the molecular dynamic simulations were subsequently confined to the region encompassing residues 160–180, using the following simulation procedure: the target temperature started at 0 K to reach the final temperature, 300 K, within 18 ps. After 30 ps of stabilization at 300 K, the molecular dynamic phase lasted 100 ps at 300 K, with a time step of 0.001 ps and a dielectric constant ( $\epsilon$ ) of 4.0. The conformations trapped at 300 K were visualized by using the VMD (Visual Molecular Dynamics) program (Humphrey *et al.*, 1996). The mean of the conformations, which was subjected to 500 steps of energy minimization, was used in structure comparison.

## Results

### *Production and purification of the mutant enzymes*

Site-directed mutagenesis was used to replace the amino acid residues located at positions 164, 170, 171, 179 and 237 in PER-1 by those found at the same positions in TEM-1 having no significant activity against expanded-spectrum cephalosporins. Thus, the single mutants A164R, H170N, A171E, N179D and T237A, and also two double mutants, A164R+A171E and A164R+N179D, were constructed. In addition, Lys242 in PER-1, which could be the counterpart of the lysine residue found at position 240 in various TEM-ESBLs (Bush and Jacoby, 1997), was replaced by a glutamic acid residue, as found in TEM-1 at position 240 (Sutcliffe, 1978). Finally, Arg220 in PER-1, which is equivalent to Arg244 located on the  $\beta$ 4 strand in TEM-1, was replaced by a leucine. SDS-PAGE analysis of crude extracts showed that all the mutant  $\beta$ -lactamases except three, were expressed in normal amounts (data not shown). Indeed, when compared with the wild-type enzyme, the A164R, N179D and A164R+N179D mutants were expressed at very low levels and the various purification attempts carried out in order to determine the kinetic features of the three mutants remained unsuccessful. Production and purification of the other enzymes, which were all active, were performed as described previously (Bouthors *et al.*, 1998).

Isoelectric focusing, carried out on the purified enzymes, indicated that three mutants, A164R+A171E, H170N and T237A, displayed  $pI$  values indistinguishable from that of the wild-type protein ( $pI = 5.4$ ). Conversely, the  $pI$  values found for the R220L, A171E and K242E mutants were shifted towards more acidic values ( $pI = 5.2, 5.0$  and  $4.9$ , respectively) (data not shown). Finally, the isoelectric points of the three mutants A164R, N179D and A164R+N179D could not be determined since the corresponding crude extracts contained no significant  $\beta$ -lactamase activity.

### *Kinetic analysis*

The steady-state kinetic parameters  $k_{cat}$  and  $K_m$  for penicillin G, cephalothin, cefotaxime (CTX), ceftazidime (CAZ) and aztreonam (AZT) were determined from the purified active  $\beta$ -lactamases. The values obtained are shown in Table II.

**Wild-type  $\beta$ -lactamase.** As expected, the values of the rate constants obtained for the wild-type PER-1  $\beta$ -lactamase were similar to those reported previously (Bouthors *et al.*, 1998). The enzyme was characterized by a high apparent affinity for penicillin G, cephalothin and AZT ( $K_m$  values ranging from 23 to 147  $\mu$ M), but a poor apparent affinity for the expanded-spectrum cephalosporins CTX and CAZ (441 and 4150  $\mu$ M, respectively). Conversely, the  $k_{cat}$  values for the last two drugs (41 and 109  $s^{-1}$ , respectively) were markedly higher than those

found for the other  $\beta$ -lactam antibiotics ( $k_{cat}$  values ranging between 8 and 11  $s^{-1}$ ) (Table II).

**Mutants of residues located in the  $\Omega$ -loop.** Four positions were investigated at the level of the  $\Omega$ -loop region of the protein: 164, 170, 171 and 179. Four single mutants (A164R, H170N, A171E and N179D) and two double mutants (A164R + A171E and A164R + N179D) were analyzed, as described below.

**Mutants A164R, A171E and A164R+A171E.** As observed previously (Bouthors *et al.*, 1998), no significant enzymatic activity was detected with the A164R mutant. By contrast, the substitution of the alanine residue found at position 171 in PER-1 by a glutamate resulted in no significant modifications of the  $k_{cat}$  and  $K_m$  values, when compared with the wild-type enzyme (Table II). Similarly, the double mutant A164R+A171E yielded an active enzyme which showed  $k_{cat}$  and  $K_m$  values similar to those of PER-1, but the  $k_{cat}/K_m$  ratios for CTX, CAZ and AZT were increased by at least an order of magnitude (Table II).

**Mutants N179D and A164R+N179D.** Position 179 is well conserved in class A  $\beta$ -lactamases, where an aspartate residue is generally found (Table III). The mutation Asn179  $\rightarrow$  Asp in PER-1, either in the N179D mutant or in the double mutant A164R + N179D, resulted in a complete loss of activity and the corresponding enzymes could not be purified.

**Mutant H170N.** A histidine residue is found at position 170 in PER-1, instead of the highly conserved Asn170 found in most of the class A  $\beta$ -lactamases described so far (Ambler *et al.*, 1991) (Table III). For penicillin G, cephalothin, CAZ and AZT, the H170N mutant displayed  $k_{cat}$  and  $K_m$  values similar to those of PER-1 (Table II). By contrast, a marked increase in  $k_{cat}$  was observed for CTX (5.5-fold) with a concomitant decrease in the apparent affinity ( $\sim 3$ -fold), thus resulting in a 2-fold increase in  $k_{cat}/K_m$ .

**Mutants of residues located in the  $\alpha/\beta$  domain.** Three positions were studied in the  $\alpha/\beta$  domain: position 220, position 237 on strand  $\beta$ 3 and position 242 on the loop connecting  $\beta$ 3 and  $\beta$ 4 (Bouthors *et al.*, 1998).

**Mutant R220L.** An arginine is found at position 220 in PER-1 (Table III). This residue might be the equivalent of Arg244 in the TEM enzymes, as previously suggested (Matagne and Frere, 1995). Replacement of Arg220 by a leucine yielded a mutant (R220L) displaying no significant modifications of the kinetic parameters for penicillin G and cephalothin. For the other drugs (CTX, CAZ and AZT), a general increase in apparent affinity was observed. In addition, a significant decrease in  $k_{cat}$  for CTX was noticed (3.2-fold) (Table II).

**Mutant T237A.** As in the ESBLs TEM-5 and TEM-24 (Sougakoff *et al.*, 1989; Chanal *et al.*, 1992), position 237, which contributes to the oxyanion pocket and corresponds to an alanine in TEM-1, is occupied by a threonine in PER-1 (Table III). The replacement of Thr237 by Ala yielded an enzyme which exhibited a higher apparent affinity for most of the substrates tested, particularly for CTX and AZT ( $K_m$  values lowered by 40- and 10-fold, respectively). By contrast, specific and divergent variations of  $k_{cat}$  were observed for CAZ (6-fold increase) and CTX (4-fold decrease), but, overall, the  $k_{cat}/K_m$  ratio for all the substrates tested was markedly increased.

**Mutant K242E.** Lysine 242, which would be located in PER-1 on a large loop connecting strands  $\beta$ 3 and  $\beta$ 4 (Bouthors *et al.*, 1998), could be the counterpart of Lys240 found in

**Table II.** Kinetic parameters<sup>a</sup> for hydrolysis of  $\beta$ -lactam antibiotics by PER-1 and the corresponding mutants

Enzyme	Penicillin G			Cephalothin			Cefotaxime			Ceftazidime			Aztreonam		
	$K_m$	$k_{cat}$	$k_{cat}/K_m$	$K_m$	$k_{cat}$	$k_{cat}/K_m$	$K_m$	$k_{cat}$	$k_{cat}/K_m$	$K_m$	$k_{cat}$	$k_{cat}/K_m$	$K_m$	$k_{cat}$	$k_{cat}/K_m$
PER-1	27	8	296	23	8	348	441	41	93	4150	109	26	147	11	75
	$\pm 3$	$\pm 0.2$	$\pm 22$	$\pm 1$	$\pm 0.1$	$\pm 17$	$\pm 42$	$\pm 2$	$\pm 8$	$\pm 611$	$\pm 15$	$\pm 8$	$\pm 15$	$\pm 0.7$	$\pm 15$
A164R	ND <sup>b</sup>	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
A171E	49	9	184	47	14	298	284	39	137	4309	134	31	42	5	119
	$\pm 2$	$\pm 0.2$	$\pm 13$	$\pm 2$	$\pm 0.2$	$\pm 20$	$\pm 6$	$\pm 0.4$	$\pm 4$	$\pm 95$	$\pm 3$	$\pm 1$	$\pm 1$	$\pm 0.1$	$\pm 5$
N179D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
A164R+A171E	23	9	391	27	12	444	300	61	203	2087	123	59	45	15	333
	$\pm 0.01$	$\pm 0.01$	$\pm 0.2$	$\pm 1$	$\pm 0.2$	$\pm 25$	$\pm 19$	$\pm 2$	$\pm 20$	$\pm 207$	$\pm 10$	$\pm 11$	$\pm 4$	$\pm 0.5$	$\pm 39$
A164R+N179D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
H170N	15	4	266	10	4	400	1286	225	175	4051	150	37	78	14	179
	$\pm 0.7$	$\pm 0.04$	$\pm 15$	$\pm 0.4$	$\pm 0.05$	$\pm 9$	$\pm 85$	$\pm 11$	$\pm 8$	$\pm 616$	$\pm 19$	$\pm 10$	$\pm 9$	$\pm 1$	$\pm 19$
R220L	15	10	666	10	4	400	167	13	78	1329	83	62	39	6	154
	$\pm 0.4$	$\pm 0.06$	$\pm 47$	$\pm 0.3$	$\pm 0.03$	$\pm 14$	$\pm 8$	$\pm 0.3$	$\pm 5$	$\pm 33$	$\pm 1$	$\pm 3$	$\pm 3$	$\pm 0.1$	$\pm 14$
T237A	22	11	500	7	15	2143	10	10	1000	2181	678	311	14	18	1286
	$\pm 0.4$	$\pm 0.05$	$\pm 11$	$\pm 0.3$	$\pm 0.1$	$\pm 117$	$\pm 0.7$	$\pm 0.2$	$\pm 94$	$\pm 232$	$\pm 56$	$\pm 59$	$\pm 1$	$\pm 0.2$	$\pm 110$
K242E	34	8	235	36	12	333	873	94	108	3672	147	40	104	16	154
	$\pm 0.5$	$\pm 0.03$	$\pm 4$	$\pm 0.6$	$\pm 0.06$	$\pm 7$	$\pm 0.7$	$\pm 0.05$	$\pm 0.1$	$\pm 31$	$\pm 1$	$\pm 0.6$	$\pm 5$	$\pm 0.4$	$\pm 11$

<sup>a</sup>Units for  $K_m$ ,  $k_{cat}$  and  $k_{cat}/K_m$  are  $\mu\text{M}$ ,  $\text{s}^{-1}$  and  $\text{s}^{-1}.\text{mM}^{-1}$ , respectively. Values of standard errors are indicated below the kinetic values.

<sup>b</sup>ND, not detectable ( $k_{cat} < 0.05 \text{ s}^{-1}$ ).

**Table III.** Multiple sequence alignment of  $\beta$ -lactamases PER-1, TEM-1, TEM-5, SHV-8, *Streptomyces albus* G and *Staphylococcus aureus* PC1

$\beta$ -Lactamase	Amino acid at position <sup>a</sup>							Reference
	164	170	171	179	220	237	240	
PER-1	A	H	A	N <sup>b</sup>	R	T	K <sup>c</sup>	Nordmann and Naas (1994)
TEM-1	R	N	E	D	L	A	E	Sutcliffe (1978)
TEM-5	S	N	E	D	L	T	K	Sougakoff <i>et al.</i> (1989)
SHV-8	R	N	E	N	L	A	E	Rashced <i>et al.</i> (1997)
<i>S.albus</i> G	R	N	S	D	R	Q	R	Dehottay <i>et al.</i> (1987)
<i>S.aureus</i> PC1	R	N	Y	D	L	A	I	East and Dyck (1989)

<sup>a</sup>Numbering according to Ambler *et al.* (1991).

<sup>b</sup>Boldface letters indicate the amino acids shared by PER-1 and other enzymes.

<sup>c</sup>Residue number 242 in PER-1.

various ESBLs displaying a high activity against CAZ and AZT (Bush and Jacoby, 1997) (Table III). This residue was replaced by a glutamic acid, which is the residue found at position 240 in TEM-1 (Table III). As shown in Table II, the steady-state kinetic parameters determined from the K242E mutant were nearly identical with those measured from PER-1.

## Discussion

We have described the catalytic behaviour of various PER-1 mutants in which residues 164, 170, 171, 179, 220, 237 and 242 were modified.

In PER-1, an alanine residue is found at position 164 instead of the highly conserved arginine identified in the other class A  $\beta$ -lactamases (Ambler *et al.*, 1991). As reported above and as observed previously (Bouthors *et al.*, 1998), replacement of Ala164 by Arg in PER-1 resulted in a mutant protein which could not be detected on SDS-PAGE analysis and which displayed no detectable  $\beta$ -lactamase activity. In order to explain such a result, theoretical three-dimensional models of the class A  $\beta$ -lactamase PER-1 and the corresponding mutant A164R

were constructed and compared with each other (Figure 1A). Despite the relatively low degree of identity found at the amino acid level between PER-1 and the other class A  $\beta$ -lactamases, homology modelling was used to generate the model structures of PER-1 and the A164R enzyme because it is now well established that class A  $\beta$ -lactamases form a super family of enzymes that are all characterized by a very similar structural organization, particularly at the level of the active site (Joris *et al.*, 1991). Molecular dynamic simulations were then performed from the models in order to assess the extent of the conformational modifications that could occur in the  $\Omega$ -loop region of the mutant by comparison with that of the wild-type enzyme. Based on the results obtained, the  $\Omega$ -loop region in PER-1 appears to be characterized by fairly high flexibility (data not shown). Such a result could be related to the fact that the PER-1  $\Omega$ -loop is not stabilized by several ionic-bonding interactions, thus contrasting with the four salt bridges found in TEM-1 between the  $\Omega$ -loop residues Arg161 and Asp163, Arg164 and Glu171, Arg164 and Asp179, and Asp176 and Arg178 (Jelsch *et al.*, 1993). Therefore, it is likely that the Ala164  $\rightarrow$  Arg substitution induces in PER-1 significant conformational modifications at the level of the  $\Omega$ -loop. Accordingly, the topology of the main-chain atoms between residues 171–179 is significantly different in the A164R mutant, when compared with PER-1 (r.m.s.d. = 0.6 Å) (Figure 1A). In the the wild-type enzyme model, the  $\Omega$ -loop conformation is generally wider than in the mutant structure, the side chain of Asp172 being oriented outwards the loop. By contrast, in the A164R enzyme, the bulky side chain of Arg164 would point inwards the  $\Omega$ -loop and, due to a putative salt bridge bonding interaction, the Asp172 side chain would be reoriented towards that of Arg164 (Figure 1A). Such a salt bridge cannot be established without a significant conformational modification of the 172 region (Figure 1A), which accounts for the instability and the loss of activity of the mutant enzyme.

Contrasting with the behaviour of the A164R mutant, the substitution in PER-1 of Ala171 by a glutamic acid yielded an enzyme characterized by kinetic parameters very similar to

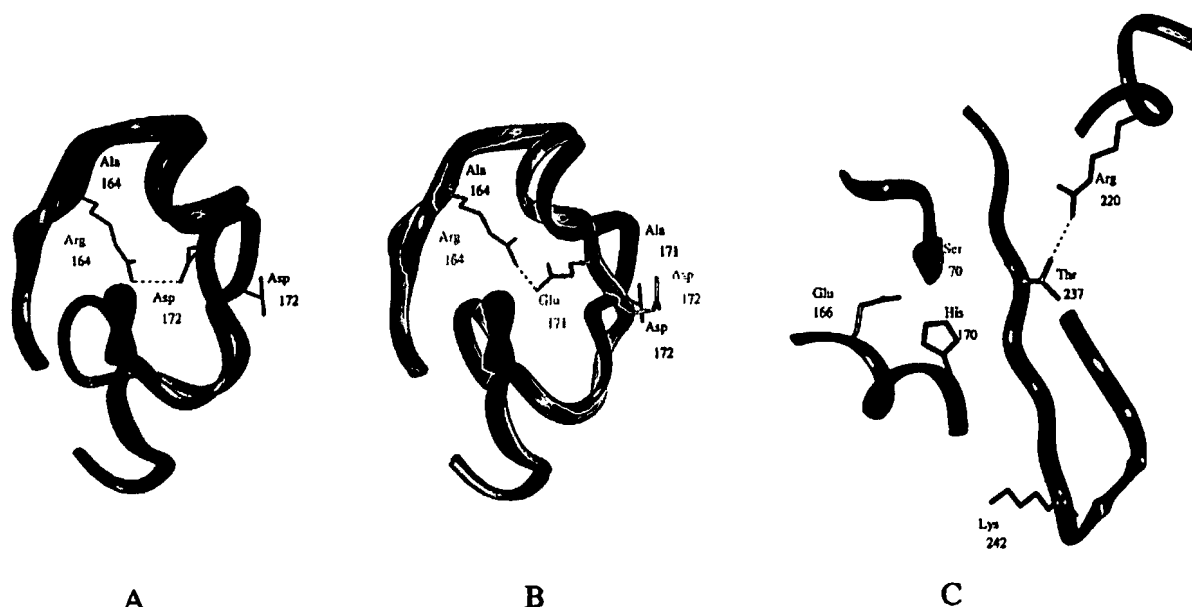


Fig. 1. Ribbon models of PER-1 (blue) and the two mutants A164R (red) and A164R+A171E (yellow). (A) Superposition of PER-1 and A164R; (B) superposition of PER-1 and A164R+A171E; (C) model of strands  $\beta$ 3 and  $\beta$ 4 in wild-type PER-1. In (A) and (B), the models represent the mean of the conformations obtained by dynamic simulations of the 160–180 region encompassing the  $\Omega$ -loop (residues 161–179). Model (C) was obtained by homology modelling and energy minimization, as described in the Materials and methods section. Thin colour sticks: side chains of residues 70, 164, 166, 170, 171, 172, 220 and 237. Green dotted lines: hydrogen bonds. The figures were generated by using the program InsightII from Molecular Simulations.

those obtained from the wild-type enzyme. Moreover, it is noteworthy that introduction of the Ala171  $\rightarrow$  Glu mutation in the inactive mutant A164R yielded a double mutant, A164R+A171E, which was fully active and exhibited  $k_{\text{cat}}$  and  $K_{\text{m}}$  values nearly identical with those of PER-1 and A171E (Table II). Strikingly, in the minimized mean of the A164R+A171E structure model obtained by molecular dynamic simulations, the side chain of Glu171 could be adequately oriented to be hydrogen bonded and/or to form a strong ionic bond with the side chain of Arg164, while that of Asp172 would be consequently oriented outwards the  $\Omega$ -loop, i.e. in a position nearly identical with that found in the wild-type PER-1 enzyme (Figure 1B), explaining why the double mutant remained fully active despite the presence of Arg164.

Two other amino acids (Asn179 and His170) were investigated in the  $\Omega$ -loop of PER-1. The asparagine residue, found at position 179 in PER-1 as in the ESBL SHV-8 (Rasheed *et al.*, 1997) (Table III), was initially thought to play a specific role in the activity of PER-1 against expanded-spectrum cephalosporins. Unexpectedly, replacement of Asn179 by an aspartate, which is a residue conserved in a large number of class A  $\beta$ -lactamases, was highly deleterious for the overall  $\beta$ -lactamase activity of the two PER mutants N179D and A164R+N179D (Table II). It must be pointed out that the interaction between residues 164 and 179, which links the two ends of the  $\Omega$ -loop region in class A  $\beta$ -lactamases, is important for a suitable positioning of the key catalytic residue Glu166 (Knox, 1995; Matagne *et al.*, 1998). Therefore, it is tempting to speculate that the presence of an aspartate residue at position 179 in the inactive mutants N179D and A164R+N179D could alter significantly the position of Glu166 and, thereby, the  $\beta$ -lactamase activity.

The histidine found at position 170 in PER-1 corresponds to a highly conserved asparagine residue in the other class A

$\beta$ -lactamases (Ambler *et al.*, 1991) (Table III). Unexpectedly, the kinetic parameters exhibited by the H170N mutant were similar to those obtained from PER-1, except for a 5.5-fold increase in the  $k_{\text{cat}}$  value for CTX with a concomitant decrease in the apparent affinity for this antibiotic. Palzkill *et al.* (1994) have reported that the replacement of the highly conserved Asn170 by a histidine in TEM-1 yielded an active enzyme showing unmodified catalytic constants. Taken altogether, these data suggest that His170 is not a key residue for the substrate profile of PER-1 and one can hypothesize that this residue was present in the ancestor of the PER-1  $\beta$ -lactamase and has been conserved during the evolution process leading to PER-1.

Three positions in PER-1 were investigated in the region of the  $\alpha/\beta$  domain forming one of the two edges of the active site. Residue 237, located on the  $\beta$ 3 strand, belongs to the so-called oxyanion pocket and is involved in the binding of  $\beta$ -lactams (Ghuysen, 1994; Matagne *et al.*, 1998). In PER-1, a threonine is found at position 237 (Figure 1C), which is located between the KTG triad and Ser238. Strikingly, it has been previously reported that various TEM-type ESBLs harbour a A237T substitution (Bush and Jacoby, 1997). Moreover, another hydroxylated residue (a serine) is found naturally at position 237 in the class A  $\beta$ -lactamase from *Proteus vulgaris* which displays a high catalytic activity against CTX and it has been shown that the substitution Ser237  $\rightarrow$  Ala in this enzyme leads to a decrease in the catalytic efficiency against this drug (Tamaki *et al.*, 1994). Therefore, the decrease in  $k_{\text{cat}}$  observed for CTX with the T237A mutant of PER-1 confirms that Thr237 is important for the catalytic activity of PER-1 towards this drug. However, the general increase in  $k_{\text{cat}}/K_{\text{m}}$  observed for the T237A mutant of PER-1 against CTX, CAZ and AZT, which is due to a general increase in apparent affinity towards cephalosporins, was rather unexpected (Table II). Nonetheless, these results were confirmed by modifying the arginine found at position 220 in PER-1. Indeed,

according to the hypothetical model of PER-1 shown in Figure 1C, the side chain of Arg220 would point towards the active site cavity and could be hydrogen-bonded to that of Thr237. As a consequence of this structural organization, it is likely that both residues contribute to adjusting the topology of the oxyanion pocket, as previously suggested for other class A  $\beta$ -lactamases (Matagne and Frere, 1995). In accordance with such a model, the replacement of Arg220 by Leu in PER-1, which leads to the loss of the hydrogen-bonding interactions between residues 220 and 237, yielded a mutant enzyme (R220L) showing kinetic properties similar to those exhibited by the T237A mutant, i.e. a significant decrease in the catalytic activity against CTX associated with a better apparent affinity for expanded-spectrum cephalosporins and AZT (see Table II).

Finally, we also studied the lysine residue found at position 242 at the end of the  $\beta$ 3 strand in PER-1 (Figure 1C), which might be the counterpart of Lys240 found in various TEM-type ESBLs (Bush and Jacoby, 1997). The replacement of Lys242 in PER-1 by a glutamic acid residue, which is the residue found at position 240 in TEM-1 (Table III), yielded a mutant enzyme with kinetic properties very similar to those of PER-1. This result indicates that Lys242 does not play in PER-1 a role equivalent to that of the lysine found at position 240 in the TEM-type ESBLs.

In conclusion, PER-1 is a class A ESBL which illustrates well the fact that enzymes showing a high level of divergence in their amino acid sequences can share very similar substrate profiles. Furthermore, our results indicate that, in contrast to the TEM-type ESBLs, the PER-1 activity towards expanded-spectrum cephalosporins does not stem from the presence in the active site of a limited number of residues having a specific role in the hydrolysis of these drugs. The X-ray structure determination of PER-1, which is in progress, will aid further understanding of the structure-activity relationships of this peculiar class A  $\beta$ -lactamase.

## Acknowledgements

We thank N.Dagoncau-Blanchard for technical assistance. We are indebted to Professor P.Nordmann and to Dr T.Naas for the gift of the recombinant plasmid pRAZ1 encoding *bla*<sub>PER-1</sub>. This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM, grant CRI950601). A.T.Bouthors is a Fellow of the Institut SmithKline-Beecham (SmithKline-Beecham Pharmaceuticals, Nanterre, France).

## References

- Ambler, R.P., Coulson, F.W., Frere, J.M., Ghuysen, J.M., Joris, B., Forsman, M., Levesque, R.C., Tiraby, G. and Walcy, S.G. (1991) *Biochem. J.*, **276**, 269–272.
- Birnboim, H.C. and Doly, J. (1979) *Nucleic Acids Res.*, **7**, 1513.
- Bouthors, A.T., Dagoncau-Blanchard, N., Naas, T., Nordmann, P., Jarlier, V. and Sougakoff, W. (1998) *Biochem. J.*, **330**, 1443–1449.
- Brunger, A.T. (1988) In Isaac, N.W. and Taylor, M.R. (eds), *Crystallographic Computing 4: Techniques and Technologies*. Clarendon Press, Oxford, pp. 126–140.
- Bush, K. and Jacoby, G.A. (1997) *J. Antimicrob. Chemother.*, **39**, 1–3.
- Chanal, C., Poupert, M.C., Sirot, D., Labia, R., Sirot, J. and Cluzel, R. (1992) *Antimicrob. Agents Chemother.*, **36**, 1817–1820.
- Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 2172–2175.
- Cornish-Bowden, A. (1995) *Analysis of Enzyme Kinetic Data*. Oxford University Press, Oxford.
- Dehottay, P., Dusart, J., De Meester, F., Joris, B., Van Becumen, J., Erpicum, T., Frere, J.M. and Ghuysen, J.M. (1987) *Eur. J. Biochem.*, **166**, 345–350.
- East, A.K. and Dyke, G.H. (1989) *J. Gen. Microbiol.*, **135**, 1001–1015.
- Ghuysen, J.M. (1994) *Trends Microbiol.*, **2**, 327–380.
- Herzberg, O. and Moul, J. (1987) *Science*, **236**, 694–701.
- Humphrey, W., Dalka, A. and Schulten, K. (1996) *J. Mol. Graphics*, **14**, 33–38.
- Jacob-Dubuisson, F., Lamotte-Brasseur, J., Dideberg, O., Joris, B. and Frere, J.M. (1991) *Protein Engng*, **4**, 811–819.
- Jacoby, J.A. and Medeiros, A.A. (1991) *Antimicrob. Agents Chemother.*, **35**, 1697–1704.
- Jelsch, C., Mourcy, L., Masson, J.M. and Samama, J.P. (1993) *Proteins: Struct. Funct. Genet.*, **16**, 364–383.
- Jonis, B., Ledent, P., Dideberg, O., Fonze, E., Lamotte-Brasseur, J., Kelly, J.A., Ghuysen, J.M. and Frere, J.M. (1991) *Antimicrob. Agents Chemother.*, **35**, 2294–2301.
- Knox, J.R. (1995) *Antimicrob. Agents Chemother.*, **39**, 2593–2601.
- Kunkel, T.A., Roberts, J.D. and Zakow, R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
- Laemmli, U.K. (1970) *Nature*, **227**, 680–685.
- Lorber, B. and Giegé, R. (1992) In Ducruix, A. and Giegé, R. (eds), *Crystallization of Nucleic Acids and proteins: a Practical Approach*. IRL Press, Oxford, p. 29.
- Matagne, A. and Frere, J.M. (1995) *Biochim. Biophys. Acta*, **1246**, 109–127.
- Matagne, A., Lamotte-Brasseur, J. and Frere, J.M. (1998) *Biochem. J.*, **330**, 581–598.
- McClary, J.A., Witney, F. and Geisselsoder, J. (1989) *BioTechniques*, **3**, 282–289.
- Messing, J. (1983) *Methods Enzymol.*, **101**, 20–78.
- Moews, P.C., Knox, J.R., Dideberg, O., Charlier, P. and Frere, J.M. (1990) *Proteins: Struct. Funct. Genet.*, **7**, 156–171.
- Nordmann, P. and Naas, T. (1994) *Antimicrob. Agents Chemother.*, **38**, 104–114.
- Nordmann, P., Ronco, E., Naas, T., Dupont, C., Michel-Briand, Y. and Labia, R. (1993) *Antimicrob. Agents Chemother.*, **37**, 962–969.
- O'Callaghan, C.H., Morris, A., Kirby, S.M. and Shingler, A.M. (1972) *Antimicrob. Agents Chemother.*, **1**, 283–288.
- Palzkill, T., Quyen-Quyen, L., Venkatachalam, K.V., LaRocco, M. and Ocra, H. (1994) *Mol. Microbiol.*, **12**, 217–229.
- Peitsch, M.C. (1996) *Biochem. Soc. Trans.*, **24**, 274–279.
- Pridmore, R.D. (1987) *Gene*, **56**, 309–312.
- Rasheed, J.K., Jay, C., Metchock, B., Berkowitz, F., Weigel, L., Crellin, J., Steward, C., Hill, B., Medeiros, A.A. and Tenover, F.C. (1997) *Antimicrob. Agents Chemother.*, **41**, 647–653.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Sougakoff, W., Petit, A., Goussard, S., Sirot, D., Bure, A. and Courvalin, P. (1989) *Gene*, **78**, 339–348.
- Strynadka, N.C.J., Adachi, H., Jensen, S.E., Johns, K., Siclecki, A., Betzel, C., Sutton, K. and James, M.N.J. (1992) *Nature*, **359**, 700–705.
- Sutcliffe, J.G. (1978) *Proc. Natl Acad. Sci. USA*, **75**, 3737–3741.
- Tamaki, M., Nukaga, M. and Sawai, T. (1994) *Biochemistry*, **33**, 10200–10206.

Received July 23, 1998; revised January 4, 1999, accepted January 21, 1999

# Compilation of vertebrate-encoded transcription factors

Steffen Faisst and Silke Meyer

Oncologie Moléculaire, Institut Pasteur de Lille, 1 rue Calmette, BP 245, F-59019 Lille Cédex, France

Received October 16, 1991; Revised and Accepted December 2, 1991

## INTRODUCTION

Since the discovery of DNA elements controlling the initiation of transcription by RNA polymerase II some ten years ago, it has become evident that the frequency of transcription initiation depends on proteins interacting with specific DNA elements of gene regulatory regions. Especially during the last three years an enormous number of such proteins, called transcription factors, have been isolated and characterized.

In the following table, we present a listing of transcription factors. It may serve as a dictionary of transcription factors, and should help to identify putative regulatory DNA elements of not yet analysed promoter regions. To keep the listing manageable, it was limited to vertebrate-encoded factors regulating the expression of genes transcribed into mRNA, i.e. by RNA polymerase II. An alphabetical order of the listing was chosen, since otherwise most of the well characterized factors had to be placed into more than one of the listed categories, such as protein families (e.g. zinc finger proteins) or factor families (e.g. steroid hormone receptor superfamily). Names of synonyms or of homologues derived from other species are listed in the second column. In the third column, the specific regulatory DNA element (if possible, the consensus sequence) that is recognized by the respective factor, is given (please note, that also the complementary sequence represents a specific binding site, since most transcription factors act in an orientation-independent manner).

In the following columns, structural features, tissue specificity and some general informations of each factor are noted. Unfortunately, the broad range of this listing left little space to describe all important features of well characterized factors such as AP1, NF $\kappa$ B or Sp1. Also, we are sure to have missed some important elements. Hence, the following table, and especially the 'features' column, represents a rather personal and selective view, which might restrict its usefulness for some purposes. Hence, in order to avoid misinterpretations, readers are encouraged to refer to the original publications cited in the table, and to references therein.

## ACKNOWLEDGEMENTS

We are grateful to J.J. Cornelis, K.F. MacLeod and V. Laudet for helpful discussions and critical reading of the manuscript, and to J. Rommelaere for his hospitality during the elaboration of this work. S.F. is a fellow of the Deutsche Forschungsgemeinschaft, S.M. of the Bundesministerium für Forschung und Technik.

## REFERENCES

1. Nakamura, T., Donovan, D.M., Hamada, K., Sax, C.M., Norman, B., Flanagan, J.R., Ozato, K., Westphal, H. and Piatigorsky, J. (1990) *Mol. Cell. Biol.* **10**, 3700-3708.
2. Sax, C.M., Klement, J.F. and Piatigorsky, J. (1990) *Mol. Cell. Biol.* **10**, 6813-6816.
3. Decker, T., Lew, D.J., Mirkowitch, J. and Darnell, J.E. (1991) *EMBO J.* **10**, 927-932.
4. Hoffman, E.C., Reyes, H., Chu, F.-F., Sander, F., Conley, H., Brooks, B.A. and Hankinson, O. (1991) *Science* **252**, 954-958.
5. Paulson, K.E., Darnell, J.E., Rushmore, T. and Pickett, C.B. (1990) *Mol. Cell. Biol.* **10**, 1841-1852.
6. Ron, D., Brasier, M.R. and Habener, J.F. (1991) *Mol. Cell. Biol.* **11**, 2887-2895.
7. Herbst, R.S., Boczek, E.M., Darnell, J.E. and Babiss, L.E. (1990) *Mol. Cell. Biol.* **10**, 3896-3905.
8. Wasylyk, C., Wasylyk, B., Heidecker, G., Huleihel, M. and Rapp, U.R. (1989) *Mol. Cell. Biol.* **9**, 2247-2250.
9. Yamaguchi, Y., Satake, M. and Ito, Y. (1989) *J. Virol.* **63**, 1040-1048.
10. Saffen, D.W., Cole, A.J., Worley, P.F., Christy, B.A., Ryder, K. and Baraban, J.M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7795-7799.
11. Bartel, D.P., Sheng, M., Lau, L.F. and Greenberg, M.E. (1989) *Genes Dev.* **4**, 304-313.
12. Piette, J., Hirai, S.-I. and Yaniv, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3401-3405.
13. Angel, P., Smeal, T., Meek, J. and Karin, M. (1989) *New Biologist* **1**, 35-43.
14. Yang-Yen, H.-F., Chiu, R. and Karin, M. (1990) *New Biologist* **2**, 351-361.
15. Gius, D., Cao, X., Rauscher, F.J., Cohen, D.R., Curban, T. and Sukhatme, V.P. (1990) *Mol. Cell. Biol.* **10**, 4243-4255.
16. Boyle, W.J., Smeal, T., Defize, L.H.K., Angel, P., Woodgett, G.R., Karin, M. and Hunter, T. (1991) *Cell* **64**, 573-584.
17. Auwerx, J. and Sassone-Corsi, P. (1991) *Cell* **64**, 983-993.
18. Williams, T. and Tjian, R. (1991) *Genes Dev.* **5**, 670-682.
19. Mitchell, P.J., Timmons, P.M., Hebert, J.M., Rigby, P.W.J. and Tjian, R. (1991) *Genes Dev.* **5**, 105-119.
20. Courtois, S.J., Lafontaine, D.A., Lemaigre, F.P., Durvieux, S.M. and Rousseau, G.G. (1990) *Nucleic Acids Res.* **18**, 57-64.
21. Comb, M. and Goodman, H.M. (1990) *Nucleic Acids Res.* **18**, 3975-3982.
22. Winning, R.S., Shea, L.J., Marcus, S.J. and Sargent, T.D. (1991) *Nucleic Acids Res.* **19**, 3709-3714.
23. Mitchell, P.J., Wang, C. and Tjian, R. (1987) *Cell* **50**, 847-861.
24. Mercurio, F. and Karin, M. (1989) *EMBO J.* **8**, 1455-1460.
25. Emmel, E.A., Verweij, C.L., Durand, D.B., Higgins, K.M., Lacy, E., Crabtree, G.R. (1989) *Science* **246**, 1617-1620.
26. Mermond, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell* **58**, 741-753.
27. Hu, Y.F., Lüscher, B., Admon, A., Mermond, M. and Tjian, R. (1990) *Genes Dev.* **4**, 1741-1752.
28. Tilley, W.D., Marcelli, M., Wilson, J.D. and McPhaul, M.J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 327-331.
29. Forman, B.M. and Samuels, H.H. (1990) *Mol. Endocrinology* **4**, 1293-1301.
30. Ladias, J.A.A. and Karathanasis, S.K. (1991) *Science* **251**, 561-565.
31. Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M. and Ishii, S. (1989) *EMBO J.* **8**, 2023-2028.
32. Hai, T., Liu, F., Coukos, W.J. and Green, M.R. (1989) *Genes Dev.* **3**, 2083-2090.

33. Hurst, H.C., Masson, M., Jones, N.C. and Lee, K.A.W. (1990) *Mol. Cell. Biol.* **10**, 6192–6203.
34. Kara, C.J., Liou, H.-C., Ivashkiv, L.B. and Glimcher, L.H. (1990) *Mol. Cell. Biol.* **10**, 1347–1357.
35. Rooney, R.J., Raychaudhuri, P. and Nevins, R. (1990) *Mol. Cell. Biol.* **10**, 5138–5149.
36. Tan, T.-H., Horikoshi, M. and Roeder, R.G. (1989) *Mol. Cell. Biol.* **9**, 1733–1745.
37. Park, R.E., Haseltine, W.A. and Rosen, C.A. (1988) *Oncogene* **3**, 275–279.
38. Ivashkiv, L.B., Liou, H.-C., Kara, C.J., Lamph, W.W., Venna, I.M. and Glimcher, L.H. (1990) *Mol. Cell. Biol.* **10**, 1609–1621.
39. Lee, T.-C., Chow, K.-L., Fang, P. and Schwartz, R.J. (1991) *Mol. Cell. Biol.* **11**, 5090–5100.
40. Cortes, P., Buckbinder, L., Leza, M.A., Rak, N., Hearing, P., Merino, A. and Reinberg, D. (1988) *Genes Dev.* **2**, 975–990.
41. Tassios, P.T. and La Tangué, N.B. (1990) *New Biologist* **2**, 1123–1134.
42. Wada, T., Watanabe, H., Usada, Y. and Handa, H. (1991) *J. Virol.* **65**, 557–564.
43. Sheng, M., Thompson, M.A. and Greenberg, M.E. (1991) *Science* **252**, 1427–1430.
44. Lewis, C.D., Clark, S.P., Felsenfeld, G. and Gould, H. (1988) *Genes Dev.* **2**, 863–873.
45. Barberis, A., Windenhorn, K., Vitelli, L. and Busslinger, M. (1990) *Genes Dev.* **4**, 849–859.
46. Boxer, L.M., Prywes, R., Roeder, R.G. and Kedes, L. (1989) *Mol. Cell. Biol.* **9**, 515–522.
47. Walsh, K. (1989) *Mol. Cell. Biol.* **9**, 2191–2201.
48. Cai, M. and Davis, R.W. (1990) *Cell* **61**, 437–446.
49. Brahm, R.J. and Kornberg, R.D. (1987) *Mol. Cell. Biol.* **7**, 403–409.
50. Moncollin, V., Stalder, R., Verdier, J.-M., Sentenac, A., Egly, J.-M. (1990) *Nucleic Acids Res.* **18**, 4817–4823.
51. Baker, R.E. and Masison, D.C. (1990) *Mol. Cell. Biol.* **10**, 2458–2467.
52. Fraser, J.D., Irving, B.A., Crabtree, G.R. and Weiss, A. (1991) *Science* **251**, 313–316.
53. Thalmeier, K., Synovzik, H., Mertz, R., Winnacker, E.-L. and Lipp, M. (1989) *Genes Dev.* **3**, 527–536.
54. Umek, R.M., Friedman, A.D. and McKnight, S.L. (1991) *Science* **251**, 248–251.
55. Pei, D. and Shih, C. (1990) *J. Virol.* **64**, 1517–1522.
56. Park, E.A., Roesler, W.J., Liu, J., Klemm, D.J., Gurney, A.L., Thatcher, J.D., Shuman, J., Friedman, A. and Hanson, R.W. (1990) *Mol. Cell. Biol.* **10**, 6264–6272.
57. McKnight, S.L., Lane, M.D. and Gluecksohn-Waelsch, S. (1989) *Genes Dev.* **3**, 2021–2024.
58. Riggs, K.J., Merrell, K.T., Wilson, G. and Calame, K. (1991) *Mol. Cell. Biol.* **11**, 1765–1769.
59. Tsai, S.Y., Sagami, I., Wang, H., Tsai, M. and O'Malley, B.W. (1987) *Cell* **50**, 701–709.
60. de Verneuil, H. and Metzger, D. (1990) *Nucleic Acids Res.* **18**, 4489–4497.
61. Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) *Cell* **53**, 11–24.
62. Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarente, L. and Sharp, P.A. (1988) *Cell* **53**, 25–35.
63. Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) *Cell* **50**, 863–872.
64. Dutta, A., Stoeckle, M.Y. and Hanafusa, H. (1990) *Genes Dev.* **4**, 243–254.
65. Hooft van Huijsduijnen, R., Li, X.Y., Black, D., Matthes, H., Benoist, C. and Mathis, D. (1990) *EMBO J.* **9**, 3119–3127.
66. Lobanenko, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V. and Goodwin, G.H. (1990) *Oncogene* **5**, 1743–1753.
67. Wuari, J. and Schibler, U. (1990) *Cell* **63**, 1257–1266.
68. Mueller, C.R., Maire, P. and Schibler, U. (1990) *Cell* **61**, 279–291.
69. Murre, C., Schonleber, McCaw, P. and Baltimore, D. (1989) *Cell* **56**, 777–783.
70. Murre, C., Voronova, A. and Baltimore, D. (1991) *Mol. Cell. Biol.* **11**, 1156–1160.
71. Nelson, C., Shen, L., Meister, A., Fodor, E. and Rutter, W.J. (1990) *Genes Dev.* **4**, 1035–1043.
72. Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. and Cleary, M.L. (1990) *Cell* **60**, 535–545.
73. Ruezinsky, D., Beckmann, H. and Kadesh, T. (1991) *Genes Dev.* **5**, 29–37.
74. Kamps, M.P., Look, A.T. and Baltimore, D. (1991) *Genes Dev.* **5**, 358–368.
75. Schlissel, M., Voronova, A. and Baltimore, D. (1991) *Genes Dev.* **5**, 1367–1376.
76. Kenny, S. and Guntaka, R.V. (1990) *Virology* **176**, 483–493.
77. Bagchi, S., Raychaudhuri, P. and Nevins, J.R. (1990) *Cell* **62**, 659–669.
78. Chellappan, S.P., Hiebert, S., Mudryj, M., Horowitz, J.M. and Nevins, J.R. (1991) *Cell* **65**, 1053–1061.
79. Mudryj, M., Hiebert, S.W. and Nevins, J.R. (1990) *EMBO J.* **9**, 2179–2184.
80. Neill, S.D., Hemstrom, C., Virtanen, A. and Nevins, J.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2008–2012.
81. Neill, S.D. and Nevins, J.R. (1991) *J. Virol.* **65**, 5364–5373.
82. Mudryj, M., Devoto, S.H., Hiebert, S.W., Hunter, T., Pines, J. and Nevins, J.R. (1991) *Cell* **65**, 1243–1253.
83. Bagchi, S., Weinmann, R. and Raychaudhuri, P. (1991) *Cell* **65**, 1063–1072.
84. Leza, M.A. and Hearing, P. (1988) *J. Virol.* **62**, 3003–3013.
85. Watanabe, H., Wada, T. and Handa, H. (1990) *EMBO J.* **9**, 841–847.
86. Wu, F.K., Garcia, J.A., Harrich, D. and Gaynor, R.B. (1988) *EMBO J.* **7**, 2117–2129.
87. Clark, L., Pollock, R.M. and Hay, R.T. (1988) *Genes Dev.* **2**, 991–1002.
88. Bruder, J.T. and Hearing, P. (1989) *Mol. Cell. Biol.* **9**, 5143–5153.
89. Ostapchuk, P., Scheirle, G. and Hearing, P. (1989) *Mol. Cell. Biol.* **9**, 2787–2797.
90. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P. (1990) *Mol. Cell. Biol.* **10**, 3456–3467.
91. Cao, X., Koski, R.A., Gashler, A., McKiernan, M., Morris, C.F., Gaffney, R., Hay, R.V. and Sukhatme, V.P. (1990) *Mol. Cell. Biol.* **10**, 1931–1939.
92. Christy, B. and Nathans, D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8737–8741.
93. Suva, L.J., Ernst, M. and Rodan, G.A. (1991) *Mol. Cell. Biol.* **11**, 2503–2510.
94. Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Duboule, D. and Charnay, P. (1990) *EMBO J.* **9**, 1209–1218.
95. Joseph, L.J., Le Beau, M.M., Jamieson, G.A., Acharya, S., Shows, T.B., Rowley, J.D. and Sukhatme, V.P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7164–7168.
96. Tsukiyama, T., Niwa, O. and Yokoro, K. (1989) *Mol. Cell. Biol.* **9**, 4670–4676.
97. Gaub, M.-P., Bellard, M., Scheuer, I., Chambon, P., Sassone-Corsi, P. (1990) *Cell* **63**, 1267–1276.
98. Fawell, S.E., Lees, J.A., White, R. and Parker, M.G. (1990) *Cell* **60**, 953–962.
99. Fawell, S.E., Lees, J.A., White, R. and Parker, M.G. (1990) *Cell* **60**, 953–962.
100. Kageyama, R. and Pastan, I. (1989) *Cell* **59**, 815–825.
101. Kageyama, R., Merlino, G.T. and Pastan, I. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5016–5020.
102. Pogonec, P., Boulukos, K.E. and Ghysdael, J. (1989) *Oncogene* **4**, 643–653.
103. Wasylyk, B., Wasylyk, C., Flores, P., Bègue, A., Leprince, D. and Stehelin, D. (1990) *Nature* **346**, 191–193.
104. Ho, I.-C., Bhat, N.K., Gottschalk, L.R., Lindsten, R., Thompson, C.B., Papas, T.S. and Leiden, J.M. (1990) *Science* **250**, 814–818.
105. Bhat, N.K., Thompson, C.B., Lindsten, T., June, C.H., Fujiwara, S., Koizumi, S., Fisher, R.J. and Papas, T.S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3723–3727.
106. Gunther, C.V., Nye, J.A., Bryner, R.S. and Graves, B.J. (1990) *Genes Dev.* **4**, 667–679.
107. Wasylyk, C., Gutman, A., Nicholson, R. and Wasylyk, B. (1991) *EMBO J.* **10**, 1127–1134.
108. Lee, C.Q., Yun, Y., Hoeffler, J.P. and Habener, J.F. (1990) *EMBO J.* **9**, 4455–4465.
109. Decker, T., Lew, D.J. and Darnell, J.E. (1991) *Mol. Cell. Biol.* **11**, 5147–5153.
110. Bamhart, K.M., Kim, C.G. and Sheffery, M. (1989) *Mol. Cell. Biol.* **9**, 2606–2614.
111. Orkin, S.H. (1990) *Cell* **63**, 665–672.
112. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F., D'Agati, V., Orkin, S.H. and Constantini, F. (1991) *Nature* **349**, 257–260.
113. Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.H. and Engel, J.D. (1990) *Genes Dev.* **4**, 1650–1662.
114. Evans, T. and Felsenfeld, G. (1991) *Mol. Cell. Biol.* **11**, 843–853.
115. Hannon, R., Evans, T., Felsenfeld, G. and Gould, H. (1991) *Proc. Natl. Acad. Sci. USA* **86**, 3004–3008.
116. Ho, I.-C., Vorhees, P., Marin, N., Oakley, B.K., Tsai, S.-F., Orkin, S.H. and Leiden, J.M. (1991) *EMBO J.* **10**, 1187–1192.
117. Ko, L.J., Yamamoto, M., Leonard, M.W., George, K.M., Ting, P. and Engel, J.D. (1991) *Mol. Cell. Biol.* **11**, 2778–2784.
118. Joulin, V., Bories, D., Eléouet, J.-F., Labastie, M.-C., Chrétien, S., Mattéi, M.-G. and Roméo, P.-H. (1991) *EMBO J.* **10**, 1809–1816.
119. Bodner, M., Castrillo, J.-L., Theill, L.E., Deemick, T., Ellisman, M. and Karin, M. (1988) *Cell* **55**, 505–518.

120. Ingraham, H.A., Flynn, S.E., Voss, J.W., Albert, V.R., Kapiloff, M.S., Wilson, L. and Rosenfeld, M.G. (1990) *Cell* 61, 1021-1033.
121. Schaufele, F., West, B.L. and Reudelhuber, T. (1990) *Nucleic Acids Res.* 18, 5235-5244.
122. Jonat, C., Rahmsdorf, H.J., Park, K.-K., Cato, A.C.B., Gebel, S., Ponta, H. and Herrlich, P. (1990) *Cell* 62, 1189-1204.
123. Gallinari, P., La Bella, F. and Heintz, N. (1989) *Mol. Cell. Biol.* 9, 1566-1575.
124. Hamada, K., Gleason, S., Levi, B.-Z., Hirschfeld, S., Apella, E. and Ozato, K. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8289-8293.
125. Lenardo, M., Rustgi, A.K., Schievella, A.R., and Bernards, R. (1989) *EMBO J.* 8, 3351-3355.
126. Gilmore, T.D. (1990) *Cell* 62, 841-843.
127. Henseling, U., Schmidt, W., Scholer, H.R., Gruss, P. and Hatzopoulos, A.K. (1990) *Mol. Cell. Biol.* 10, 4100-4109.
128. Rustgi, A.K., Van'T Veer, L.J. and Bernards, R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8707-8710.
129. Dailey, L., Roberts, S.B. and Heintz, N. (1988) *Genes Dev.* 2, 1700-1712.
130. Majello, B., Arcone, R., Toniatti, C. and Ciliberto, G. (1990) *EMBO J.* 9, 457-465.
131. Van Wijnen, A.J., Wright, K.L., Massung, R.F., Gerretsen, M., Stein, J.L. and Stein, G.S. (1988) *Nucleic Acids Res.* 16, 571-592.
132. Means, A.L. and Farnham, P.J. (1990) *Mol. Cell. Biol.* 10, 653-661.
133. Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W. and Azizkhan, J.C. (1990) *Mol. Cell. Biol.* 10, 6632-6641.
134. Böhnlein, E., Lowenthal, J.W., Siekevitz, M., Ballard, D.W., Franza, B.R. and Greene, W.C. (1988) *Cell* 53, 827-836.
135. Franza, B.R., Josephs, S.F., Gilman, M.Z., Ryan, W. and Clarkson, B. (1987) *Nature* 330, 391-395.
136. Cereghini, S., Yaniv, M. and Cortese, R. (1990) *EMBO J.* 9, 2257-2263.
137. Baumhueter, S., Mendel, D.B., Conley, P.B., Kuo, C.J., Turk, C., Graves, M.K., Edwards, C.A., Courtois, G. and Crabtree, G.R. (1990) *Genes Dev.* 4, 372-379.
138. Kuo, C.J., Conley, P.B., Hsieh, C.-L., Francke, U. and Crabtree, S. (1990) *Proc. Natl. Acad. Sci. USA* 87, 9838-9842.
139. Toniatti, C., Demartis, A., Monaci, P., Nicosia, A. and Ciliberto, G. (1990) *EMBO J.* 9, 4467-4475.
140. De Simone, V., De Magistris, L., Lazzaro, D., Gerstner, J., Monaci, P., Nicosia, A. and Cortese, R. (1991) *EMBO J.* 10, 1435-1443.
141. Rey-Campos, J., Chouard, T., Yaniv, M. and Cereghini, S. (1991) *EMBO J.* 10, 1445-1457.
142. Mendel, D.B., Hansen, L.P., Graves, M.K., Conley, P.B. and Crabtree, G.R. (1991) *Genes Dev.* 5, 1042-1056.
143. Bach, I., Mattei, M.-G., Cereghini, S. and Yaniv, M. (1991) *Nucleic Acids Res.* 19, 3553-3559.
144. Lai, E., Prezioso, V.R., Smith, E., Litvin, O., Costa, R.H. and Darnell, J.E. (1990) *Genes Dev.* 4, 1427-1436.
145. Lai, E., Prezioso, V.R., Tao, W., Chen, W.S. and Darnell, J.E. (1991) *Genes Dev.* 5, 416-427.
146. Sladek, F.M., Zhong, W., Lai, E. and Darnell, J.E. (1990) *Genes Dev.* 4, 2353-2365.
147. Costa, R.H., Grayson, D.R. and Darnell, J.R. (1989) *Mol. Cell. Biol.* 9, 1415-1425.
148. Grange, T., Roux, J., Rigaud, G. and Pictet, R. (1991) *Nucleic Acids Res.* 19, 131-139.
149. Abravaya, K., Phillips, B. and Morimoto, R.I. (1991) *Mol. Cell. Biol.* 11, 586-592.
150. Sorger, P.K. (1991) *Cell* 65, 363-366.
151. Goldenberg, C.J., Luo, Y., Fenna, M., Baler, R., Weinmann, R. and Voellmy, R. (1988) *J. Biol. Chem.* 263, 19734-19739.
152. Cunniff, N.F.A., Wagner, J. and Morgan, W.D. (1991) *Mol. Cell. Biol.* 11, 3504-3514.
153. Whelan, J., Cordle, S.R., Henderson, E., Weil, P.A. and Stein, R. (1990) *Mol. Cell. Biol.* 10, 1564-1572.
154. Leshkowitz, D. and Walker, M.D. (1991) *Mol. Cell. Biol.* 11, 1547-1552.
155. Roman, C., Platero, J.S., Shuman, J. and Clamann, K. (1990) *Genes Dev.* 4, 1404-1415.
156. Peterson, C.L. and Calame, K. (1989) *Mol. Cell. Biol.* 9, 776-786.
157. Yan, C. Tamm, I. (1991) *Proc. Natl. Acad. Sci. USA* 88, 144-148.
158. de Groot, R.P., Auwerx, J., Karperien, M., Staels, B. and Krujer, W. (1991) *Nucleic Acids Res.* 19, 775-781.
159. Imam, A.M.A., Ackrill, A.M., Dale, T.C., Kerr, I.M. and Stark, G.R. (1990) *Nucleic Acids Res.* 18, 6573-6580.
160. Harada, H., Willison, K., Sakakibara, J., Miyamoto, M., Fujita, T. and Taniguchi, T. (1990) *Cell* 63, 303-312.
161. Yu-Lee, L.-Y., Hrachovy, J.A., Stevens, A.M. and Schwarz, L.A. (1990) *Mol. Cell. Biol.* 10, 3087-3094.
162. Pine, R., Decker, T., Kessler, D.S., Levy, D.E. and Darnell, J.E. (1990) *Mol. Cell. Biol.* 10, 2448-2457.
163. Kessler, D.S., Veals, S.A., Fu, X.-X. and Levy, D.E. (1990) *Genes Dev.* 4, 1753-1765.
164. Gosset, L.A., Kelvin, D.J., Sternberg, E.H. and Olson, E.L. (1989) *Mol. Cell. Biol.* 9, 5022-5033.
165. Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M.B., Kourilsky, P., Baeuerle, B.A. and Israel, A. (1990) *Cell* 62, 1007-1018.
166. Leask, A., Rosenberg, M., Vassar, R. and Fuchs, E. (1990) *Genes Dev.* 4, 1985-1998.
167. Ramji, D.P., Tadros, M.H., Hardon, E.M. and Cortese, R. (1991) *Nucleic Acids Res.* 19, 1139-1146.
168. Carlsson, P., Erikson, P. and Bjursell, G. (1990) *Gene* 94, 295-301.
169. Huang, H.-C., Sundseth, R. and Hansen, U. (1990) *Genes Dev.* 4, 287-298.
170. Lo, K., Landau, N.R. and Smale, S.T. (1991) *Mol. Cell. Biol.* 11, 5229-5243.
171. Imbert, J., Zafarullah, M., Culotta, V.C., Gedamu, L. and Hamer, D. (1989) *Mol. Cell. Biol.* 9, 5315-5323.
172. Baldwin, A.S., LeClair, K.P., Singh, H. and Sharp, P.A. (1990) *Mol. Cell. Biol.* 10, 1406-1414.
173. Fan, C.-M. and Maniatis, T. (1990) *Genes Dev.* 4, 29-42.
174. Maekawa, T., Sakura, H., Sudo, T. and Ishii, S. (1989) *J. Biol. Chem.* 264, 14591-14593.
175. Mar, J.H. and Ordahl, C.P. (1990) *Mol. Cell. Biol.* 10, 4271-4283.
176. Cserjesi, P. and Olson, E.N. (1991) *Mol. Cell. Biol.* 11, 4854-4862.
177. Labbé, S., Prévost, J., Remondelli, P., Leone, A. and Séguin, C. (1991) *Nucleic Acids Res.* 19, 4225-4231.
178. Arriza, J.L., Weinberger, C., Cerelli, G., Glaser, T.M., Handelin, B.L., Housman, D.E. and Evans, R.M. (1987) *Science* 237, 268-275.
179. Culotta, V.C. and Hamer, D.H. (1989) *Mol. Cell. Biol.* 9, 1376-1380.
180. Mueller, P.R., Salser, S.J. and Wold, B. (1988) *Genes Dev.* 2, 412-427.
181. Parisi, M.A. and Clayton, D.A. (1991) *Science* 252, 965-969.
182. Ness, S.A., Marknell, A. and Graf, T. (1989) *Cell* 59, 1115-1125.
183. Lüscher, B. and Eisenman, R.N. (1990) *Genes Dev.* 4, 2235-2241.
184. Biedenkapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.-H. (1988) *Nature* 335, 835-837.
185. Howe, K.M., Reakes, C.F.L. and Watson, R.J. (1990) *EMBO J.* 9, 161-169.
186. Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) *Science* 250, 1149-1151.
187. Ariga, H., Imamura, Y. and Iguchi-Ariga, S.M.M. (1989) *EMBO J.* 8, 4273-4279.
188. Dang, C.V., Barrett, J., Villa-Garcia, M., Resar, L.M.S., Kato, G.J. and Fearon, E.R. (1991) *Mol. Cell. Biol.* 11, 954-962.
189. Blackwood, E.M. and Eisenman, R.N. (1991) *Science* 251, 1211-1217.
190. Prendergast, G.C., Lawe, D. and Ziff, E.B. (1991) *Cell* 65, 395-407.
191. Vaidya, T.B., Rhodes, S.J., Taparowsky, E.J. and Konieczny, S.F. (1989) *Mol. Cell. Biol.* 9, 3576-3579.
192. Davis, R.L., Cheng, P., Lassar, A.B. and Weintraub, H. (1990) *Cell* 60, 733-746.
193. Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P., Weintraub, H. and Lassar, A.B. (1988) *Science* 242, 405-411.
194. Lewin, B. (1991) *Cell* 64, 303-312.
195. Molitor, J.A., Walker, W.H., Doerre, S., Ballard, D.W. and Greene, W.C. (1990) *Proc. Natl. Acad. Sci. USA* 87, 10028-10032.
196. Ballard, D.W., Walker, W.H., Doerre, S., Sista, P., Molitor, J.A., Dixon, E.P., Pfeffer, N.J., Hannik, M. and Greene, W.C. (1990) *Cell* 63, 803-814.
197. Shirakawa, F., Chedid, M., Suttles, J., Pollok, B.A. and Mizel, S.B. (1989) *Mol. Cell. Biol.* 9, 959-964.
198. Nolan, G.P., Gosh, S., Liou, H.-C., Tempst, P. and Baltimore, D. (1991) *Cell* 64, 961-969.
199. Urban, M.B., Schreck, R. and Baeuerle, P.A. (1991) *EMBO J.* 10, 1817-1825.
200. Beckmann, H., Su, L.-K. and Kadesch, T. (1990) *Genes Dev.* 4, 167-179.
201. Peterson, C.L., Eaton, S. and Calame, K. (1988) *Mol. Cell. Biol.* 8, 4972-4980.
202. Kober, M., Reith, W., Herrero-Sanchez, C. and Mach, B. (1990) *Mol. Cell. Biol.* 10, 965-971.
203. Dorn, A., Benoist, C. and Mathis, D. (1989) *Mol. Cell. Biol.* 9, 312-320.
204. Watson, M.A. and Milbrandt, J. (1989) *Mol. Cell. Biol.* 9, 4213-4219.
205. Fahrner, T.J., Carroll, S.L. and Milbrandt, J. (1990) *Mol. Cell. Biol.* 10, 6454-6459.
206. Evans, M.J. and Scarpulla, R.C. (1990) *Genes Dev.* 4, 1023-1034.

207. Yoza, B.K. and Roeder, R.G. (1990) *Mol. Cell. Biol.* **10**, 2145–2153.
208. Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) *EMBO J.* **9**, 1883–1888.
209. Poellinger, L. and Roeder, R.G. (1989) *Mol. Cell. Biol.* **9**, 747–756.
210. Kamps, M.P., Corcoran, L., Lebowitz, J.H. and Baltimore, D. (1990) *Mol. Cell. Biol.* **10**, 5464–5472.
211. Mul, Y.M., Verrijzer, C.P. and van der Vliet, P.C. (1990) *J. Virol.* **64**, 5510–5518.
212. Roberts, S.B., Segil, N. and Heintz, N. (1991) *Science* **253**, 1022–1026.
213. Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and de Crombrughe, B. (1988) *Cell* **52**, 405–414.
214. Goyal, N., Knox, J. and Gronostajski, R.M. (1990) *Mol. Cell. Biol.* **10**, 1041–1048.
215. Gounari, F., De Francesco, R., Schmitt, J., van der Vliet, P., Cortese, R. and Stunnenberg, H. (1990) *EMBO J.* **9**, 559–566.
216. Mermond, N., Williams, T.J. and Tjian, R. (1988) *Nature* **332**, 557–561.
217. Randak, C., Brabletz, T., Hergenröther, M., Sobotta, I. and Serfling, E. (1990) *EMBO J.* **9**, 2529–2536.
218. Fiering, S., Northrop, J.P., Nolan, G.P., Mattila, P.S., Crabtree, G.R. and Herzenberg, L.A. (1990) *Genes Dev.* **4**, 1823–1834.
219. Flanagan, W.M., Corthesy, B., Bram, R.J. and Crabtree, G.R. (1991) *Nature* **352**, 803–807.
220. Mignotte, V., Wall, L., deBoer, E., Grosveld, F. and Romeo, P.-H. (1989) *Nucleic Acids Res.* **17**, 37–54.
221. Talbot, D. and Grosveld, F. (1991) *EMBO J.* **10**, 1391–1398.
222. Caruso, M., Iacobini, C., Passananti, C., Felsani, A. and Amati, P. (1990) *EMBO J.* **9**, 947–955.
223. Yu, Y.-T. and Nadal-Ginard, B. (1989) *Mol. Cell. Biol.* **9**, 1839–1849.
224. Shannon, M.F., Gamble, J.R. and Vadas, M.A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 674–678.
225. Poli, V., Mancini, F.P. and Cortese, R. (1990) *Cell* **63**, 643–653.
226. Descombes, P., Chokier, M., Lichtsteiner, S., Falvey, E. and Schibler, U. (1990) *Genes Dev.* **4**, 1541–1551.
227. Isshiki, H., Akira, S., Sugita, T., Nishio, Y., Hashimoto, S., Pawlowski, T., Suematsu, S. and Kishimoto, T. (1991) *New Biologist* **3**, 63–70.
228. Chang, C.-J., Chen, T.-T., Lei, H.-Y., Chen, D.-S. and Lee, S.-C. (1990) *Mol. Cell. Biol.* **10**, 6642–6653.
229. Williams, P., Ratajczak, T., Lee, S.C. and Ringold, G.M. (1991) *Mol. Cell. Biol.* **11**, 4959–4965.
230. Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1988) *EMBO J.* **7**, 4221–4229.
231. Wirth, T., Priess, A., Annweiler, A., Zwilling, S. and Oeler, B. (1991) *Nucleic Acids Res.* **19**, 43–51.
232. Müller, C.L., Feldhaus, A.L., Rooney, J.W., Rhodes, L.D., Sibley, C.H. and Singh, H. (1991) *Mol. Cell. Biol.* **11**, 4885–4894.
233. Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) *Cell* **60**, 461–472.
234. Rosner, M.H., De Santo, R.J., Arnheiter, A. and Staudt, L.M. (1991) *Cell* **64**, 1103–1110.
235. Schöler, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990) *EMBO J.* **9**, 2185–2195.
236. Meijer, D., Graus, A., Kraay, R., Langeveld, A., Mulder, M.P. and Grosveld, G. (1990) *Nucleic Acids Res.* **18**, 7357–7365.
237. Suzuki, N., Rohdewohld, H., Neuman, T., Gruss, P. and Schöler, H.R. (1990) *EMBO J.* **9**, 3723–3732.
238. Schreiber, E., Harschman, K., Kemler, I., Malipiero, U., Schaffner, W. and Fontana, A. (1990) *Nucleic Acids Res.* **18**, 5495–5503.
239. Smith, D.P. and Old, R.W. (1991) *Nucleic Acids Res.* **19**, 815–821.
240. Chalepakidis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M. and Gruss, P. (1991) *Cell* **66**, 873–884.
241. Kakkis, E., Riggs, K.J., Gillespie, W. and Calame, K. (1989) *Nature* **339**, 718–721.
242. Schüle, R., Umesono, K., Mangelsdorf, D.J., Bolado, J., Pike, J.W. and Evans, R.M. (1990) *Cell* **61**, 497–504.
243. Rorth, P., Nerlov, C., Blasi, F. and Morton, J. (1990) *Nucleic Acids Res.* **18**, 5009–5021.
244. Wasyslyk, C., Flores, P., Gutman, A. and Wasyslyk, B. (1989) *EMBO J.* **8**, 3371–3378.
245. Gutman, A. and Wasyslyk, B. (1990) *EMBO J.* **9**, 2241–2246.
246. Issemann, I. and Green, S. (1990) *Nature* **347**, 645–650.
247. DeMarzo, A.M., Beck, C.A., Onate, S.A. and Edwards, D.P. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 72–76.
248. Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H. and Chambon, P. (1990) *EMBO J.* **9**, 1603–1614.
249. Keller, A.D. and Maniatis, T. (1991) *Genes Dev.* **5**, 868–879.
250. McCormick, A., Brady, H., Fukushima, J. and Karin, M. (1991) *Genes Dev.* **5**, 1490–1503.
251. Cockell, M., Stevenson, B.J., Strubin, M., Hagenbüchle, O. and Wellauer, P.K. (1989) *Mol. Cell. Biol.* **9**, 2464–2476.
252. Roux, E., Strubin, M., Hagenbüchle, O. and Wellauer, P.K. (1989) *Genes Dev.* **3**, 1613–1624.
253. Petruccio, S., Wellauer, P.K. and Hagenbüchle, O. (1990) *Mol. Cell. Biol.* **10**, 254–264.
254. Goebel, M.G., Moreau-Gachelin, F., Ray, D., Tambourin, P., Tavittian, A., et al. (1990) *Cell* **61**, 1165–1166.
255. Paul, R., Schuetze, S., Kozak, S.L., Kozak, C.A. and Kabat, D. (1991) *J. Virol.* **65**, 464–467.
256. Postel, E.H., Mango, S.E. and Flint, S.J. (1989) *Mol. Cell. Biol.* **9**, 5123–5133.
257. Glass, C.K., Devary, O.V. and Rosenfeld, M.G. (1990) *Cell* **63**, 729–738.
258. Ellinger-Ziegelbauer, H. and Dreyer, C. (1991) *Genes Dev.* **5**, 94–104.
259. Rowe, A., Eager, N.S.C. and Brickell, P.M. (1991) *Development* **111**, 771–778.
260. Reith, W., Herrero-Sanchez, C., Kober, M., Silacci, P., Berte, C., Barras, E., Fey, S. and Mach, B. (1990) *Genes Dev.* **4**, 1528–1540.
261. Yan, D.-H. and Hung, M.-C. (1991) *Mol. Cell. Biol.* **11**, 1875–1882.
262. Thornell, A., Hallberg, B. and Grundström, T. (1988) *Mol. Cell. Biol.* **8**, 1625–1637.
263. Boral, A.L., Okenquist, S.A. and Lenz, J. (1989) *J. Virol.* **63**, 76–84.
264. Wagner, B.J., Hayes, T.E., Hoban, C.J. and Cochran, B.H. (1990) *EMBO J.* **9**, 4477–4484.
265. Schmidt, M.C., Zhou, Q. and Berk, A.J. (1989) *Mol. Cell. Biol.* **9**, 3299–3307.
266. Jackson, S.P., MacDonald, J.J., Lees-Miller, S. and Tjian, R. (1990) *Cell* **63**, 155–165.
267. Lemaigre, F.P., Lafontaine, D.A., Courtois, S.J., Durvieux, S.M. and Rousseau, G.G. (1990) *Mol. Cell. Biol.* **10**, 1811–1814.
268. Sartorelli, V., Webster, K.A. and Kedes, L. (1990) *Genes Dev.* **4**, 1811–1822.
269. Saffer, J.D., Jackson, S.P. and Thurston, S.J. (1990) *Genes Dev.* **4**, 659–666.
270. Su, W., Jackson, S., Tjian, R. and Echols, H. (1991) *Genes Dev.* **5**, 820–826.
271. Saffer, J.D., Jackson, S.P. and Annarella, M.B. (1991) *Mol. Cell. Biol.* **11**, 2189–2199.
272. Graham, R. and Gilman, M. (1991) *Science* **251**, 189–192.
273. Shaw, P.E., Schröter, H. and Nordheim, A. (1989) *Cell* **56**, 563–572.
274. Manak, J.R. and Prywes, R. (1991) *Mol. Cell. Biol.* **11**, 3652–3659.
275. Van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.* **10**, 123–132.
276. Waterman, M.L. and Jones, K. (1990) *New Biologist* **2**, 621–636.
277. Waterman, M.L., Fischer, W.H. and Jones, K.A. (1991) *Genes Dev.* **5**, 656–669.
278. Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes Dev.* **5**, 880–894.
279. Davidson, I., Xiao, J.H., Rosales, R., Staub, A. and Chambon, P. (1988) *Cell* **54**, 931–942.
280. Xiao, J.H., Davidson, I., Matthes, H., Garnier, J.-M. and Chambon, P. (1991) *Cell* **65**, 551–568.
281. Wefald, F.C., Devlin, B.H. and Williams, S. (1990) *Nature* **344**, 260–262.
282. Maldonado, E., Ha, I., Cortes, P., Weis, L. and Reinberg, D. (1990) *Mol. Cell. Biol.* **10**, 6335–6347.
283. Tainura, T., Sumita, K., Fujino, I., Aoyama, A., Horikoshi, M., Hoffmann, A., Roeder, R.G., Muramatsu, M. and Mikoshiba, K. (1991) *Nucleic Acids Res.* **19**, 3861–3865.
284. Greenblatt, J. (1991) *Cell* **66**, 1067–1070.
285. Carr, C.S. and Sharp, P.A. (1990) *Mol. Cell. Biol.* **10**, 4384–4388.
286. Ben-Levy, R., Faktor, O., Berger, I. and Shaul, Y. (1989) *Mol. Cell. Biol.* **9**, 1804–1809.
287. Goto, M., Tamura, T., Mikoshiba, K., Masamune, Y. and Nakanishi, Y. (1991) *Nucleic Acids Res.* **19**, 3959–3963.
288. Forrest, D., Hallböök, F., Persson, H. and Vennström, B. (1991) *EMBO J.* **10**, 269–275.
289. Laudet, V., Begue, A., Henry-Dutoit, C., Joubel, A., Martin, P., Stehelin, D. and Saule, S. (1991) *Nucleic Acids Res.* **19**, 1105–1112.
290. Lazar, M.A., Berrodin, T.J. and Harding, H.P. (1991) *Mol. Cell. Biol.* **11**, 5005–5015.
291. Brent, G.A., Larsen, P.R., Harney, J.W., Koenig, R.J. and Moore, D.D. (1989) *J. Biol. Chem.* **264**, 178–182.
292. Baniahmad, A., Steiner, C., Köhne, A.C. and Renkawitz, R. (1990) *Cell* **61**, 505–514.



293. Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M. and Di Lauro, R. (1990) *EMBO J.* 9, 3631–3639.
294. Avvedimento, V.E., Musti, A.M., Ueffing, M., Obici, S., Gallo, A., Sanchez, M., DeBrasi, D. and Gottesman, M.E. (1991) *Genes Dev.* 5, 22–28.
295. Mizuno, K., Gonzales F.J. and Kimura, S. (1991) *Mol. Cell. Biol.* 11, 4927–4933.
296. Gregor, P.D., Sawadogo, M. and Roeder, R.G. (1990) *Genes Dev.* 4, 1730–1740.
297. Watt, F. and Molloy, P.L. (1988) *Nucleic Acids Res.* 16, 1471–1486.
298. Kovesdi, I., Reichel, R. and Nevins, J.R. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2180–2184.
299. Workman, J.L., Roeder, R.G. Kingston, R.E. (1990) *EMBO J.* 9, 1299–1308.
300. Kaulen, H., Pognonec, P., Gregor, P.D. and Roeder, R.G. (1991) *Mol. Cell. Biol.* 11, 412–424.
301. Iyer, S.V., Davis, D.L., Seal, S.N. and Burch, B.E. (1991) *Mol. Cell. Biol.* 11, 4863–4875.
302. Baker, A.R., McDonnell, D.P., Hughes, M., Crisp, T.M., Mangelsdorf, D.J., Haussler, M.R., Pike, J.W., Shine, J. and O'Malley, B.W. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3294–3298.
303. Call, K.W., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeager, H. and Housman, D.E. (1990) *Cell* 60, 509–520.
304. Rauscher, F.J., Morris, J.F., Tournay, O.E., Cook, D.M. and Curran, T. (1990) *Science* 250, 1259–1262.
305. Buckler, A.J., Pelletier, J., Haber, D.A., Glaser, T. and Housman, D.E. (1991) *Mol. Cell. Biol.* 11, 1701–1712.
306. Madden, S.L., Cook, D.M., Morris, J.F., Gashler, A., Sukhatme, V.P. and Rauscher, F.J. (1991) *Science* 253, 1550–1553.
307. Saatcioglu, F., Perry, D.J., Pasco, D.S. and Fagan, J.B. (1990) *Mol. Cell. Biol.* 10, 6408–6416.
308. Weinrich, S.L., Meister, A. and Rutter, W.J. (1991) *Mol. Cell. Biol.* 11, 4985–4997.
309. Didier, D.K., Schiffenbauer, J., Woulfe, S.L., Zacheis, M. and Schwartz, B.D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7322–7326.
310. Akira, S., Ishiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990) *EMBO J.* 9, 1897–1906.
311. Vasios, G., Mader, S., Gold, J.D., Leid, M., Lutz, Y., Gaub, M.-P., Chambon, P. and Gudas, L. (1991) *EMBO J.* 10, 1149–1158.
312. Baldwin, A.S., Azizkhan, J.C., Jensen, D.E., Beg, A.A. and Coodly, L.R. (1991) *Mol. Cell. Biol.* 11, 4943–4951.
313. Bakker, O. and Parker, M.G. (1991) *Nucleic Acids Res.* 19, 1213–1217.
314. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. and Chambon, P. (1986) *Nature* 320, 134–139.
315. Hasegawa, S.L., Doetsch, P.W., Hamilton, K.K., Martin, A.M., Okenquist, S.A., Lenz, J. and Boss, J.M. (1991) *Nucleic Acids Res.* 19, 4915–4920.
316. Liu, F. and Green, M.R.R. (1990) *Cell* 61, 1217–1224.

### Legend to the table:

- 1) The species in which a factor has been identified is given in superscript letters (b: bovine; c: calf; ch: chicken; f: frog; h: human; ha: hamster, m: murine; r: rat; s: simian).
- 2) The names of factors listed here are synonyms of the names listed in the first column, or homologous factors found in other species, or both. The species in which the factors have been identified are given in superscript letters (cf. 1))
- 3) If possible, consensus sequences derived from several binding sites are given. Since most transcription factors act in an orientation-independent manner, also the complementary sequence represents a specific binding site (R: purine; P: pyrimidine; N: any nucleotide).
- 4) The molecular weight of the factors was determined by different methods as indicated by superscript letters (a: SDS-PAGE; b: gel filtration; c: estimated from the corresponding cDNA). If a factor exists in two distinct forms, the  $M_r$  of the two forms is shown with a slash (e.g. 45/50); if a factor exists in several forms, or if the size was not exactly determined, the range of the  $M_r$  is shown (e.g. 45 – 50); if a factor consists of two or more polypeptides, the  $M_r$  of the polypeptides is shown (e.g. 45 + 50).
- 5) bHLH: helix-loop-helix protein containing a basic domain; bZIP: leucine zipper protein containing a basic domain; FHD: fork head domain; HD: homeodomain; HSH: helix-span-helix protein; POU: POU specific domain; zinc f.: zinc finger.
- 6) Posttranslational modifications are given in brackets (Ph: phosphoprotein; O-gly: O-glycosylated).
- 7) Compounds inhibiting the action of the respective factor are given in brackets.

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
$\alpha$ A-CRYBP1 <sup>m</sup>		GGGAAATCCC	zinc f.			Implicated in lens-specific expression of the $\alpha$ A-crystallin gene. Lens-specific transactivation requires a single binding site. Related to MBP-1 and AGIE-BP1.	1, 2
AAF <sup>h</sup>		TTTCATATTACTCT			INF- $\alpha$ , INF- $\gamma$	Has characteristics very similar to those of GAF. Induction of AAF by INF- $\gamma$ does not depend on ongoing protein synthesis.	3
AHR <sup>m,r</sup>		TCCCTGAGAAGA	280 <sup>a</sup>		$\beta$ -naphthoflavone, tetrachlorodibenzo-p-dioxin, 3-methyl-cholanthrene	Binds to Xenobiotic response element (XRE). Soluble protein complex containing the ligand-binding subunit (95kDa) and the 90 kDa heat shock protein. Translocates to the nucleus after ligand binding. Member of the steroid hormone receptor superfamily.	4, 5
AGIE-BP1 <sup>r</sup>		GGTTGGGAAATCCC	>250 <sup>a</sup> zinc f.			Binding specificity is indistinguishable from NF $\kappa$ B. Related to $\alpha$ A-CRYBP1 and MBP1.	6
ANF <sup>h,m</sup>		CTTTATCTGG		ubiquitous		Negative regulating factor of albumin gene expression.	7
AP1 <sup>h,m,r,ch</sup>	PEA1 <sup>m</sup>	TGA <sup>G</sup> /C <sup>T</sup> /AA	v-Jun: 65 <sup>a</sup> c-Jun: 39 <sup>a</sup> v-Fos: 55 <sup>a</sup> /75 <sup>a</sup> c-Fos: 55 <sup>a</sup> -65 <sup>a</sup> bZIP (Ph)	ubiquitous	TPA, EGF, Ha-ras, raf, v-mos, IL-2, NGF, TGF $\beta$ , ConA, picrotoxin, Py-mL metrazole, serum, cAMP, retinoic acid, cell. transformation, brain seizure activity, membrane depolarization	Homodimer of Jun or heterodimer between members of the Jun and Fos or Jun and ATF families. Positive or negative regulating factor of various cellular and viral promoters. Fos down-regulates immediate-early gene expression via CAAT boxes. DNA-binding is inhibited by IP1 (30-40kDa) which may interact with the Jun/Fos bZIP.	8, 9, 10, 11, 12, 13, 14, 15, 16, 17

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa), <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
AP2 <sup>h,t,f</sup>		CCCAG/CN <sub>2</sub> G/C <sub>2</sub> G/C <sub>2</sub> G/C	50 <sup>a</sup> HSH	most abundant in the neural crest lineage	TPA, cAMP, retinoic acid, forskalin (SV40 TAg, DNA methylation)	Also binds to Sp1, NF-1, and SV40 T antigen binding sites. Develop- mentally regulated. Binds as a dimer to a palindromic binding site.	18, 19, 20, 21, 22
AP3 <sup>h</sup>		TGTGGAGTAAAT	48 <sup>a</sup> /57 <sup>a</sup>		(Cyclosporin A)	Relatedness of the two proteins described as AP3 is not clear. is not clear. Both proteins bind to the GTIC motif of the SV40 promoter. May be related to TEF-2.	23, 24, 25
AP4 <sup>h</sup>		C <sub>2</sub> TCAGCTG <sub>2</sub> C <sub>2</sub> TCG	48 <sup>a</sup> bHLH, bZIP			Contains multiple protein-protein interfaces to promote homodimer formation. Interacts with AP1 in regulation of SV40 gene expression	26, 27
AR1 <sup>h,m,t</sup>		AGAACAN <sub>3</sub> TGTTCT	98.9 <sup>c</sup>		androgen	Androgen receptor, member of the steroid hormone receptor superfamily.	28, 29
ARP-1 <sup>h</sup>		TGANCCTTGACCOCT	47 <sup>a</sup> zinc f.	ubiquitous		Member of the steroid hormone receptor superfamily. Heterodimerizes with COUP. May participate in regulation of lipid metabolism and cholesterol homeostasis.	30
ATF1 <sup>h,m,t</sup>	CREB <sup>h,m</sup> , NF21 <sup>h</sup> , EATF3 <sup>h</sup> , CRE-BP1 <sup>h</sup> , CRE-BP2 <sup>m</sup> , PCRE <sup>m</sup> , TREB <sup>h</sup> , Eiv <sup>ph</sup> , Tal <sup>h</sup> , ECRE <sup>m</sup>	TGACGC <sub>2</sub> T/C <sub>2</sub> A <sub>2</sub> G/A	43 <sup>a</sup> - 72 <sup>a</sup> , bZIP, (Ph)	ubiquitous (CRE-BP1 and HB16 are most abundant in brain)	cAMP, Ca <sup>++</sup>	Binds to the cAMP response element (CRE). Family of at least ten different transcription factors encoded either by different genes or generated by differential splicing. Forms homo- or hetero-dimers with members of the ATF- or Jun families. Different dimers have different DNA-binding specificities. ATF-8 binds more efficiently to an AP1- than to an ATF site.	31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 316
BGP1 <sup>ch</sup>		GGGGGGGGGGGGGGGGGG	zinc f.	erythrocytes		Different from Sp1.	44

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
BSAP <sup>h,m</sup>		GACGCAN <sup>C</sup> TTGG <sup>G</sup> A <sup>T</sup> AN <sub>3</sub> C <sup>A</sup> AG	50 <sup>a</sup>	B cells		Binds to the promoter of tissue-specific late histone genes, no binding sites in Ig or MHC class II promoters. Present at early but not at late stages of differentiation.	45
CBP <sup>h</sup>	MAPF <sup>2,eh</sup>	ACACCCAAATATGGCGAC	35 <sup>a</sup> (Ph)			Binds to muscle regulatory element. May be closely related to SRF. Binds also to serum response element, but does not activate it.	46, 47
CBP-1 <sup>h,m</sup>	CBF-1 CP1 <sup>h</sup>	TCAC <sup>G</sup> ATGATA	39 <sup>a</sup> -49 <sup>a</sup> bHLH			Heat stable, centromere (CDE I) binding protein, implicated in chromosome segregation and transcriptional activation. Similar or identical to USF.	48, 49, 50, 51
CD28 <sup>RCh</sup>		AAAGAAATTC			CD28	Activates IL-2 promoter.	52
C/EBP <sup>h,m,z,eh</sup>	EBP20 <sup>h,m,z</sup>	GTGG <sup>T</sup> A <sup>T</sup> A <sup>T</sup> AG ATTCC	42 <sup>a</sup> bZIP	most abundant in liver, but also in brain, fat, intestine, lung, and skin		Heat resistant, is presumed to be involved in energy metabolism and may have a role in regulating the balance between cell growth and differentiation. Reveals a rather loose binding specificity that includes CCAAT boxes, the enhancer core motif, and cAMP response elements.	53, 54, 55, 56, 57, 313
CF1 <sup>m</sup>		ANATGG		ubiquitous		Acts on c-myc-, IgH-, and α-actin promoters. Interacts with PCF.	58
COUP <sup>h,z,eh</sup>	Ear-3 <sup>m</sup>	GTGTCAAAGGTCA	45-50 <sup>a</sup>			Member of the steroid hormone receptor superfamily, interacts with the non DNA-binding factor S300II. Reveals similar binding properties as v-ErbA.	29, 59, 60
CP1 <sup>h</sup>	NF-Y <sup>m</sup>	C <sup>T</sup> TTN <sub>6</sub> <sup>A</sup> GA <sup>A</sup> GCCAATCANC <sup>T</sup> GT	A: 40 <sup>a</sup> ; 34 <sup>c</sup> B: 32 <sup>a</sup> ; 23 <sup>c</sup>		v-src, serum	Consists of two subunits (A and B). Binds to CCAAT boxes found in Y boxes.	61, 62, 63, 64, 65
CP2 <sup>h</sup>	NF-Y <sup>em</sup>	C <sup>T</sup> TAGC <sup>T</sup> NN <sub>3</sub> <sup>A</sup> GRCCAATCNC <sub>3</sub> <sup>G</sup> A			v-src, serum	CCAAT box binding factor	61, 62, 64, 65

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
CTCF <sup>ph</sup>		CCCTC	130 <sup>a</sup>	erythrocyte, muscle, fibroblast		Binds to three repeats of this motif spaced at five to six basepairs intervals. Adjacent sequences are required for tight binding.	66
DBP <sup>r</sup>		TCATTITGT	43 <sup>a</sup> basic domain	ubiquitous in adults (except testis), liver- enriched	(chemically induced liver regeneration)	Highest DBP levels are found at 8 p.m., not detectable during the night or in the morning. Member of the C/EBP family, but lacks a leucine zipper domain. Related to VBP.	67, 68
E2A <sup>h</sup>	E12 <sup>h</sup> =pan1m E47 <sup>h</sup> =pan2m	G <sub>1</sub> /A <sub>1</sub> CAGNTG	E12: 67,4 <sup>c</sup> E47: 67,7 <sup>c</sup> bHLH	ubiquitous (related E2 box binding factors may be myocyte or B cell-specific)		E12 and E47 are generated by differential splicing of the E2A gene. Binds as a dimer to the κE2 motif, but also to the AP4 and USF binding sites. The E2A gene is rearranged in 95% of the t(1;19) chromosomal translocations found in acute lymphatic lymphomas. The rearrangement results in the replacement of the bHLH by a HD. E2A activates IgD-to-J rearrangement. Forms heterodimers with other bHLH proteins (such as MyoD or Id).	69, 70, 71, 72, 73, 74, 75
E2B <sup>pch</sup>		TGCAAC <sub>1</sub> TAC <sub>1</sub> T				Heat-labile, acts on U3 region of the RSV-LTR	76
E2F <sup>h,1,m</sup>		TTTTG <sub>1</sub> C <sub>1</sub> G <sub>1</sub> CGG <sub>1</sub> C	54 <sup>a</sup>		E1a, serum (cAMP, retinoic acid)	Activity most abundant during S-Phase. Factor in a proliferation-dependent signal transduction pathway. Forms complexes with the retinoblastoma gene product, cyclin A and adenovirus E4 protein. Complexes are dissociated by adenovirus E1a protein. DNA-binding is inhibited by DNA methylation.	77, 78, 79, 80, 81, 82, 83
E4F <sup>h</sup>	ETF-A <sup>h</sup>	TGACGTAAC	50 <sup>a</sup> (Ph)		E1a	Binds to a site containing the cAMP response element (CRE). Does not confer cAMP- but E1a inducibility. (cf. ATF)	84, 35

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
EBTF1 <sup>h</sup>		CGGAAGTG	60 <sup>a</sup> +53 <sup>a</sup>		ras, raf	Consists of two subunits: A DNA-binding one of 60 kDa and a transactivating one of 53 kDa. Not induced by SV40, myc and fos. May be related or identical to EF-1A.	85
EBP1 <sup>h</sup>		GGGACTTTC	57 <sup>a</sup> +60 <sup>a</sup>			Binds to the κB motif and to the SV40 enhancer core. Different from NFκB, but may be related to LBP.	86, 87
EF-1A <sup>h,m,r</sup>		CGGAAGTG		ubiquitous		Binds cooperatively to two binding sites. May be related or identical to E4TF1.	88
EF-Ch <sup>m</sup>	EPh	GTTGC <sup>T</sup> /CNG <sup>G</sup> /ACAAC		present in various human cell lines		Binds as a dimer to the Py and HBV enhancer. Alteration of the spacing between the inverted repeats of the binding site decreases binding affinity.	89
EGR-1 <sup>m,r</sup>	NGFI-A <sup>m</sup> Krox-24 <sup>m</sup> Tis8, zif268 <sup>m</sup>	CGCCCC <sup>C</sup> /GGC	82+88 <sup>a</sup> zinc f. (Ph)	ubiquitous, most abundant in brain	TPA, growth factors, metrazole, picrotoxin, retinoic acid, brain seizure activity	Consists of two polypeptides, short-lived, encoded by an immediate early response gene. Has a broad role in signal transduction pathways. (cf. WT-ZFP).	90, 91, 92, 93
EGR-2 <sup>h</sup>	Krox-20 <sup>m</sup>	CGCCCCCGC	43 <sup>a</sup> zinc f.		PMA, mitogen	May be involved in the regulation of the expression of homeobox-containing genes.	94, 95
ELP <sup>m</sup>		CAAGGTCA		undiff. embr. carcinoma cells		Implicated in negative regulation of Mo-MuLV LTR.	96
ERh <sup>m,r,chl</sup>		AGGTCA <sup>N</sup> GTGACCT	h:65c zinc f.		estrogen	Estrogen receptor (steroid hormone receptor superfamily). The three spacing nucleotides are essential for the binding affinity. ER and GR binding sites are similar. May cooperate with AP.	29, 60, 97, 98, 314

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
v-ErbA <sub>h,m,ch</sub>		GTGTCAAAGGTCA	75 <sup>a</sup> zinc f. (Ph)			Member of the steroid hormone receptor superfamily. Consists of the thyroid receptor fused to the viral gag protein. Acts as antagonist to the thyroid receptor. Its oncogenic potential correlates with its ability to repress RAR action.	60, 29, 99
ETFh		CAGCCCCCGCGCAGC	120 <sup>a</sup> ; 270 <sup>b</sup>		phorbol esters	Has a rather loose binding affinity, binds to GC-rich sequences not recognized by SP1. Acts on EGF receptor gene promoter.	100, 101
Eis-1 <sub>h,m,ch</sub>		G <sub>1</sub> C <sub>4</sub> A <sub>1</sub> C <sub>1</sub> GGAA <sub>1</sub> T <sub>1</sub> G <sub>1</sub> C	54 <sup>a</sup> (68 <sup>a</sup> ) (Ph)	B cells resting T cells	c-Ha-ras, v-src, v-mos (ConA, T cell stimulation)	In chickens also a 68 kDa-form exists. Short-lived protein with high turn-over rate. DNA-binding activity is phosphorylation-dependent. Binds <i>in vitro</i> to the PEA3 motif. May be a component of the signalling network.	102, 103, 104, 105, 106, 107
F-ACT1 <sub>ch</sub>		TCGGGA		ubiquitous		Binds to the serum response element (SRE). Its binding is mutually exclusive with that of SRF. May be a negative regulator of $\alpha$ -actin gene expression.	108
GAF <sub>h</sub>		TTTCATATTACTCT			INF- $\alpha$ , INF- $\gamma$	Has characteristics very similar to AAF. Induction of GAF by INF- $\gamma$ depends on ongoing protein synthesis.	109, 3
GATA-1 <sub>h,m,ch</sub>	Ery $\Pi$ <sub>ch</sub> , GF-1 <sub>h,m</sub> , EF-1 <sub>m</sub> , NF-E1 <sub>am,ch</sub>	T <sub>1</sub> A <sub>1</sub> GATAR	GF-1: 51 <sup>a</sup> NF-E1 <sub>a</sub> : 40 <sup>a</sup> (18 <sup>a</sup> + 19 <sup>a</sup> ) zinc f.	erythroid cells megacaryocytes mast cells		May consist of two polypeptides, required for erythroid differentiation. Activates globin genes. Erythroid cell-specific transactivation requires multiple binding sites. Plays autoregulatory role in its own expression.	110, 111, 112, 113, 114, 115
GATA-2 <sub>ch</sub>	NF-E1 <sub>ch</sub>	T <sub>1</sub> A <sub>1</sub> GATAR	56 <sup>a</sup> zinc f.	erythroid cells (limited in other tissues)		Reveals tissue-specific RNA splicing. Early factor in differentiation, whereas GATA-1 and -3 are late factors.	113, 111

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
GATA-3 <sup>h</sup> , m, ch, f	NF-E1 <sup>ch</sup>	T <sub>A</sub> GATAR	h: 49 <sup>a</sup> , 48 <sup>c</sup> ch: 55 <sup>a</sup> zinc f.	T cells, brain, definitive erythrocytes		May play a regulatory role in developing chicken brain and T cell- specific gene expression. The human GATA3 is only expressed in T cells.	113, 111, 116, 117, 118
GCF <sup>h</sup>		C <sub>1</sub> GCGC <sub>1</sub> G <sup>c</sup> G <sup>c</sup> C <sub>1</sub> G <sup>c</sup> C	100 <sup>a</sup> , 91 <sup>c</sup>			Negative regulator of EGF receptor gene.	100
GHP-1 <sup>h</sup> , f	Pic-1 <sup>m, f</sup> , PUF-1	A <sub>1</sub> TATC <sub>1</sub> T <sup>c</sup> CAT	34 <sup>a</sup> POU	adult anterior pituitary (somato-, thyro-, and lacto- trophs)		Activates the growth hormone gene and the prolactin gene, which has several binding sites.	119, 120, 121
GHP-5 <sup>f</sup>		A <sub>1</sub> TATC <sub>1</sub> T <sup>c</sup> CAT		variety of cells		Activates the growth hormone gene, binds to two sites.	121
GHP-7 <sup>f</sup>		A <sub>1</sub> TATC <sub>1</sub> T <sup>c</sup> CAT		adult anterior pituitary (somato- trophs)		Activates the growth hormone gene, binds to two sites.	121
GR <sup>h</sup>		AGAACAN <sub>3</sub> TGTTCT	88 <sup>a</sup> , 94 <sup>a</sup> zinc f.		glucocorticoid	Glucocorticoid receptor, member of the steroid hormone receptor superfamily. May be negatively regulated by AP1. The same binding site is recognized by AR, MR, and PR.	29, 122
H1TF1 <sup>h</sup>		AACAACACAAA	90 <sup>a</sup> ?			Binds to AC box of the H1 histone gene. Most probably not identical to H1NF-A.	123
H1TF2 <sup>h</sup>		GCACCAATCAGCGCGC	47 <sup>a</sup>			CCAAT box binding factor of H1 histone genes. Distinct from NF1.	123
H2RUB <sup>h</sup>		TCAGGTCACAGTGACCTGA	50 <sup>a, c</sup> zinc f.			Member of the steroid hormone receptor superfamily, binds to the cAMP response element and to the ER site.	124
H2TF1 <sup>h, m, f</sup>	MBP-2 <sup>h</sup> KB3 <sup>m</sup>	TGGGGATTCCCA	48 <sup>a</sup> , 58 <sup>a</sup> zinc f.	ubiquitous	(N-myc)	Exists in two forms. Constitutively expressed. Binds to the κB motif. Implicated in MHC class II expression.	125, 126, 127, 128



Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa), <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
H4TF1 <sup>h</sup>		GGGGGAGGG	105 <sup>a</sup> , 110 <sup>a</sup> zinc f.			Acts specifically on H4 histone gene promoter, different from SP1.	129
H4TF2 <sup>h</sup>		GGTTCCTCN <sub>4</sub> CGGTCCG	65 <sup>a</sup>			Acts specifically on H4 histone gene promoter.	129
H-APF-1 <sup>h</sup>		CTGGRAA			IL-6	Cooperates with NF-IL6, is qualitatively and quantitatively changed by IL-6.	130
HNF-A <sup>h</sup>		ATTN <sub>4</sub> ATTT				Acts on human H1, H3, and H4 histone gene promoters.	131
HIP1 <sup>h</sup>		ATTCTN(1-30)GCCA				Acts on TATA-less promoters of housekeeping genes to specify the site of transcription initiation. Was suggested to be E2F.	132, 133
HIVEN86A <sup>h</sup>		TGGGGATTCCCCA	86 <sup>a</sup>	activated T cells, B lymphocytes		Binds to κB site of the IL2Rα and HIV promoters. May be closely related to NFκB.	134, 135
HNF-1 <sup>h,r,m</sup>	LF-B1 <sup>h,r</sup> , HP1, APF, HNF-1α	GTTAATNATTAAAC	92 <sup>a</sup> POU	differentiated liver, kidney, stomach, intestine, spleen		May be implicated in endodermic differentiation. Dimerizes in the absence of its DNA recognition sequence.	136, 137, 138, 139
vHNF-1 <sup>h,r,m</sup>	LF-B3 <sup>r</sup> + LF-Bu <sup>r</sup> , HNF-1β	GTTAATNATTAAAC	72 <sup>a</sup> POU	liver, lung	retinoic acid	Replaces HNF-1 in dedifferentiated cells. Exclusive expression of vHNF-1 is associated with repression of liver-specific genes. HNF-1 and vHNF-1 have the potential to interact to produce an embryologically complex pattern of gene expression.	136, 140, 141, 142, 143
HNF-3A <sup>r</sup>	HNF-3α <sup>r</sup>	TATTGA <sup>r</sup> C <sub>7</sub> T <sup>r</sup> TTA <sup>r</sup> TCG	50 <sup>a</sup> , 48.7 <sup>c</sup> FHD	liver, intestine, lung		May contribute to the differentiation of cells in internal organs.	144, 145
HNF-3B <sup>r</sup>	HNF-3β <sup>r</sup>	TATTGA <sup>r</sup> C <sub>7</sub> T <sup>r</sup> TTA <sup>r</sup> TCG	47 <sup>a</sup> , 46 <sup>c</sup> FHD	liver, intestine, lung		May contribute to the differentiation of cells in internal organs.	144, 145

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
HNF-3C	HNF-3 $\gamma$	TATTGAC $\gamma$ TTA $\gamma$ TC	42 <sup>a</sup> , 43 <sup>c</sup> PHD	liver, intestine, testis		May contribute to the differentiation of cells in internal organs.	144, 145
HNF-4 $\gamma$		G $\gamma$ TCGA $\gamma$ TA $\gamma$ AGGT $\gamma$ GC $\gamma$ TCAC $\gamma$ TC	54 <sup>a</sup> zinc f.	liver, kidney, intestine		Member of the steroid hormone receptor superfamily, does not bind to the almost identical TR site. May be identical to LF-A1.	5, 146, 147
HNF-5 $\gamma$		TC $\gamma$ A $\gamma$ TTTC $\gamma$ TC $\gamma$ TC		liver		Binding sites are located close to those of other liver-specific factors.	148
HSP $\gamma$ h,m		(NGAAN) <sub>3</sub>	h: 83 <sup>a</sup>	ubiquitous	heat shock	Does not bind prior to or after recovery from heat shock. Does not disturb binding of other factors to the same promoter. May trimerize. Each subunit thought to bind a GAA trinucleotide.	149, 150, 151, 152
IAF $\gamma$	IEF1 $\gamma$ h,a	GCCATCTGCT		pancreatic- $\beta$ -cells		Binds to insulin control element of the insulin II gene. Binds also to the USP site, where it is inactive.	153, 154
Ig/EBP1 $\gamma$ m	$\mu$ EBP-Em	A $\gamma$ TGNATTNTG $\gamma$ TAAT- A $\gamma$ GN $\gamma$ TNA $\gamma$ TC	45 <sup>a,b</sup> bZIP	ubiquitous (highest levels in early B cells)		Member of the C/EBP family. Binds to E sites in IgH gene promoter and to the RSV LTR.	155, 156
IREBF-1 $\gamma$ m		CCGGAAATCGAAACTG	35, 5 <sup>c</sup> bZIP		interferon $\alpha$ , $\beta$	Binds to interferon response element, unrelated to ISGF-1, -2, and -3.	157
IRBP $\gamma$ m		AGTCGACT	110 <sup>a</sup>		TPA, PKA	Acts on Jun family gene promoters.	158
ISGF1 $\gamma$ h	IRF-2 $\gamma$ m, ICBP	CTTTCAGTTT		differentiated cells	interferon $\beta$	Constitutively expressed repressor of INF genes.	159, 160
ISGF2 $\gamma$ h $\gamma$ r (M+G)	IRF-1 $\gamma$ m, IBP-1	CTTCTCTTT	M: 56 <sup>a</sup> ; G: 37 <sup>a</sup> (Ph)	differentiated cells	interferon $\alpha$ (M:G), interferon $\gamma$ (G), prolactin (rat), virus infection, dsRNA	M- and G-forms of ISGF2 are generated by posttranscriptional modifications. Slow interferon response factor, depends on protein synthesis. The rat homolog is an immediate-early gene in prolactin-stimulated T cells and may play a role in cell proliferation.	159, 160, 161, 162

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
ISGF3h		GCTTCAGTTT	48 <sup>a</sup> + 84 <sup>a</sup> + 91 <sup>a</sup> + 113 <sup>a</sup>		interferon α and γ (N-ethyl maleimide)	Multimeric complex, rapid interferon response factor. Translocates to the nucleus after interferon treatment.	160, 163
KBF-1h	Rel <sup>h</sup>	TGGGGATTCCCA	50 <sup>a</sup>	ubiquitous		p50 subunit of NFκB, but also acts independently. Member of the Rel family.	165
Ker1h		GCCTGCAGGC		keratinocytes		Controls keratinocyte-specific gene expression.	166
LBpm		TAAAGCCATTT				Binds to Mo-MuLV LTR, may be related to EBPI.	96
LF-A1h,c,r		TGGAC <sup>T</sup> /C <sup>T</sup> /C <sup>N</sup> /TGGCCC	c: 40 <sup>a,b</sup>	liver		Binding site is bipartite, insertion or deletion of 1 to 4 nucleotides in the spacer region does not abolish DNA-binding.	167
LFB3m		GTTAATNATTAAC	POU	epithel of endo- and mesodermal origin	retinoic acid	Is related to and forms heterodimers with HNF-1 (cf. vHNF-1).	140, 141
LIT-1h		GCGCCCTTTGGACCT	200 <sup>b</sup>			Acts on apolipoprotein B gene promoter.	168
LSpH		GGCGGNN <sub>4</sub> GGGCGG GGTTCTTTCCGCC	63 <sup>a</sup>			Binds to two different sites on the SV40 promoter with the same affinity. Unrelated to Sp1 and MTF-1. Binds to SV40 GC boxes 2 and 3 and regulates SV40 late expression.	169
LyF-1m		PPTGGGAGR	50 <sup>a</sup>	most abundant in B and T cells		May be a member of the Ets family.	170
MBF-1m		C <sub>7</sub> TAAAAAATAAC <sub>7</sub> T <sub>7</sub> C <sub>7</sub> T <sub>7</sub>		myocytes	FGF, TGFβ	Is regulated in a reciprocal manner to MEF-2.	164
MBF-1m		TGCRRC	74 <sup>a</sup>		zinc	Binds to metal response element, may be related to MTF-1.	171

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
MBP-1 <sup>h</sup>	PRDII-BF1 <sup>h</sup> HIV-EP1 <sup>h</sup>	TGGGGATTCCCCA	298c zinc f.	ubiquitous (low levels in B cells)	TPA, serum, mitogen, virus infection	Acts on MHC class II promoter, distinct from NFκB. May be regulator of cell proliferation. May be related to PRDII-BF1, HIV-EP1, αA-CRYBP1 and AGIE-BP1.	172, 173, 174
MCBFch		CATTCT		embryonic muscle tissue		Factor regulating muscle-specific expression of the cTNT and α-actin promoters.	175
MEF-2 <sup>m</sup>		C <sub>1</sub> TAA <sup>h</sup> T <sub>1</sub> AAATA <sup>h</sup> /G		myocytes	myogenin (mitogens, FGF, TGFβ)	May participate in the coordinate regulation of genes during myogenesis. Regulated in a reciprocal manner to MBF-1.	164, 176
MEP-1 <sup>h,m</sup>		TGCRNC	115 <sup>a</sup>			Binds to metal response elements (MREd ≥ MREa = MREc > MREb > MREe > MREf).	177
MR <sup>h,r</sup>		AGAACAN <sub>3</sub> TGTTCT	107 <sup>c</sup> zinc f.		mineralocorticoids glucocorticoids	Mineralocorticoid receptor, member of the steroid hormone receptor superfamily.	178, 29
MTF-1 <sup>m</sup>		TGCRNC			zinc	Binds to the metal response elements, related to Spl.	179, 180
mtTF1 <sup>h</sup>		TTAACAGTCACCCCCAAC	24.4 <sup>c</sup> HMG			Mitochondrial transcription factor. Activates mtDNA promoters.	181
Myb <sup>h,m,r,ch</sup>		T <sub>1</sub> C <sub>1</sub> AACG <sub>1</sub> T <sub>1</sub> G	v-Myb: 48 <sup>a</sup> c-Myb: 75 <sup>a</sup>	haematopoietic system, tumor cell lines		Short-lived protein with higher affinity to tandem motifs, although it binds as a monomer. Plays a critical role in cell proliferation and differentiation.	182, 183, 184, 185
Myc <sup>h,m,r,ch</sup>		CACGTG TCTCTTA	c-Myc: 64 <sup>a</sup> /67 <sup>a</sup> b/HLH/ZIP			Binds to random DNA sequences and to at least two specific DNA sequences. Functions both in transcriptional regulation and DNA replication. Binds to DNA as a homodimer, or, with altered activity, as a heterodimer with Max.	186, 187, 188, 189, 190

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
MyoDh.s.m	MEF-1m.ch	CAACTGAC	45a bHLH (Ph)	proliferating myoblasts and differentiated myotubes	(FGF, TGFβ)	Implicated in myocyte differentiation, needs thirteen amino acids of its basic region for myocyte-specific activation. Forms heterodimers with other bHLH proteins (e. g. E2A or Id)	69, 191, 192, 193
NF1h.m,r	CTFh.m	T <sub>1</sub> /CGGA/CN <sub>5</sub> . <sub>6</sub> GCCAA	52a-66a basic domain (O-glycol.)	ubiquitous	TGFβ	Family of at least six proteins. The quantity of the different forms of NF1 varies with the growth conditions of the cells, but overall binding activity remains stable. Binds as a dimer both to palindromic and to half-side sequences. Flanking sequences, length and composition of the spacer region can greatly influence the binding affinity. Interacts also with DNA polymerase to enhance DNA replication.	20, 213, 214, 211, 215, 216
NF-ATh.m	NF-IL2-Eh	GGAGGAAAAACTGTTTCAT		activated T cells	PHA (CsA, FK506)	The concentration of NF-AT must exceed a critical threshold before transcription is initiated. Consists of a cytoplasmic and a nuclear subunit. Translocation of the cytoplasmic subunit to the nucleus is inhibited by Cyclosporin A (CsA) and FK506. Could account for the immunosuppressive effect of CsA.	25, 217, 218, 219
NF-E2m		TGACTCAG		fetal liver, myeloid cells		Binds to API site, but different from API.	220, 221
NF-Dm		GATGGGG		ubiquitous		Transcription and replication factor. Acts on Py enhancer domain D.	222
NFεh.m		GTGTCACTCA				Activates embryonic myosin heavy chain promoter, different from API and ATF.	223
NF-GMa <sup>h</sup>		GA <sub>1</sub> GGA <sub>1</sub> GTT <sup>T</sup> <sub>1</sub> GCAT <sub>1</sub> C			PMA, ConA	Acts on cytokine gene promoter.	224
NF-GMb <sup>h</sup>		TCAG <sub>1</sub> /ATA			PMA, ConA	Acts on cytokine gene promoter.	224

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa) <sub>4</sub> , domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
NF-IL6 <sup>h</sup>	LAP <sup>r</sup> IL-6DBP <sup>h,r</sup> H-APF-2 <sup>h,r</sup> AGP/EBP <sup>m</sup>	T <sup>T</sup> /G <sup>T</sup> NNGNAAT/G	NF-IL6: 40 <sup>a</sup> LAP: 32 <sup>a</sup> bZIP (Ph)	ubiquitous, most abundant in liver, heart, and muscle	lipopolysaccharide, IL-1, IL-6	Member of the C/EBP family, binds with higher affinity as heterodimer with C/EBP than as homodimer. Is involved in acute phase reactions, inflammation, and haematopoiesis. AGP/EBP may play a role in the glucocorticoid induction of the α1-acid glycoprotein gene.	130, 225, 226, 227, 228, 229, 310
NFκB <sup>h,r,m</sup>	Rel <sup>h</sup>	GGGA <sup>A</sup> /CTNT <sup>T</sup> /CCC	50 <sup>a</sup> + 65 <sup>a</sup> >200 <sup>b</sup>	ubiquitous	PMA, cAMP, IL-1, lipopolysaccharide, TNF α, TNF γ	Heterodi- or tetramer. NFκB may be one factor out of a family of rel-related proteins (40 - 125 kDa). Released after stimulation from an inactive cytoplasmic complex formed with IκBα or IκBβ. Constitutive in mature B cells, inducible in immature B cells, T cells, and non- lymphoid cells. A role for NFκB in the G <sub>0</sub> -to-G <sub>1</sub> transition was suggested.	126, 194, 195, 196, 197, 198, 199, 312
NF-μE3 <sup>h,m</sup>	CD2B <sup>pm</sup>	CCCACATGAAC	42.5 <sup>a</sup> +44 <sup>a</sup> +45 <sup>a</sup>			Binds to IgH μE3 motif. Oligomer of two to four polypeptides. Probably different from TFE3.	200, 201
NF-δ <sup>h</sup>		PyGTCAGC			cAMP?	Binds to MHC class II α-chain gene promoters.	202
NF-W1 <sup>m</sup>		GTTGCATC	64 <sup>a</sup>	mature B cells		Acts on MHC class II gene promoter, different from Oct-1 and -2.	203
NF-W2 <sup>m</sup>		GTTGCATC	64 <sup>a</sup> + 20 <sup>a</sup>	ubiquitous		Acts on MHC class II gene promoter, different from Oct-1 and -2.	203
NGF1-B <sup>r</sup>	nur77 <sup>m</sup> N10 <sup>r</sup>	AGGTCATGAACCT	63 <sup>a</sup> -88 <sup>a</sup> 61 <sup>c</sup> zinc f. (Ph)		growth hormones, brain seizure activity, TPA	Member of the steroid hormone receptor superfamily.	204, 205

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
NRF-1h,s		T <sub>1</sub> CGCGCA <sup>1</sup> T <sub>1</sub> CGCGCA <sup>1</sup> G				Nuclear factor acting on genes whose products function in mitochondria. May coordinate the expression of nuclear and mitochondrial genetic systems in response to cellular energy demands.	206
NTPh		GGAACCTCCCC		ubiquitous		Interacts on heavy chain VH gene promoter with Oct-2.	207
Oct-1h,m,r	NF-A1h, NF-11h,m, OTF-1h, OBP100h, TRF, NF-1L2h	ATGCAAAAT (CTCATGA)	90 <sup>10</sup> , 110 <sup>10</sup> (XU) (Ph)	ubiquitous	TPA, P11A	Preferentially expressed in early S-phase, also stimulates DNA replication. Has no apparent activation domains. Also binds to heptamer motif when dimerized with Oct-2. Differential phosphorylation during the cell cycle.	104, 208, 209, 210, 211, 212
Oct-2h,m,r,f	NF-A2h,m, OTF-2h	ATGCAAAAT (CTCATGA)	Oct-2A: 60 <sup>a</sup> Oct-2B: 75 <sup>a</sup> POU, bZIP	lymphoid cells, intestine, testis, kidney, nervous system	TPA, PHA, IL-1 lipopolysaccharides (TGFβ)	Trimer, acts on IgH and MHC class II gene promoters. Also binds to the heptamer motif when dimerized with Oct-1. Multiple isoforms are generated by alternative splicing. Stimulates DNA replication.	209, 210, 230, 231, 232
Oct-4h,m	Oct-3h,m, Oct-5m, NF-A3h,m	ATGCT <sub>1</sub> A <sup>1</sup> AAT (TTAAATTCa)	42 <sup>a</sup> (35 <sup>a</sup> ) POU	embryonic cells, adult ovary and testis	(retinoic acid)	Most probably, Oct-3, 4, and 5 are identical proteins. Exists in a minor form of 35 kDa. Maternal factor that is also implicated in DNA replication.	233, 234, 235, 236
Oct-6m	Tst1f, SCIPT	ATGCAAAAT TAATGARAT	45 <sup>a</sup> POU	embryonic stem cells, germ cells, adult brain and nerve cells	retinoic acid	May be identical to N-Oct-3. May control events at very early and late stages of development.	236, 237, 238
Oct-Rf		ATGCAAA <sup>1</sup> C <sup>1</sup> T		ubiquitous		Binds to the motif only in the context of the H2B box.	239
Pax-1h,m		CACCGTTCCGCTCTAGATATCTC	42 <sup>a</sup> 37 <sup>9c</sup>			Mutation of this factor leads to altered DNA-binding specificity and is associated to malformations of the vertebral column in mice.	240

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
PCFm		AGAAAGGGAAAGGA		plasmocytoma		Negative regulating factor of c-myc gene expression.	241
PEA3h,m,r, ch		AGGAAG/A			TPA, EGF, serum, v-src, c-Ha-ras, v-mos, v-raf, Py-mt	The factor binding to this motif has not yet been identified (Eis-1 has been shown to bind this motif <i>in vitro</i> ). The factor is not induced by fos. Represents a primary target of signal transduction.	103, 243, 244, 245
PPARM		AGGTCA	52.4 <sup>c</sup> zinc f.	liver, kidney heart, brown adipose tissue	peroxisome proliferators	Member of the steroid hormone receptor superfamily. Important for triglyceride and cholesterol homeostasis. May play a role in the development of liver tumors.	246
PRh, ch		AGAACAN <sub>3</sub> TGTTCT	h: 94 <sup>a</sup> /120 <sup>a</sup> ch: 79 <sup>a</sup> /109 <sup>a</sup> zinc f.		progesterone (hsp90)	Progesterone receptor, member of the steroid hormone superfamily. Exists in two forms with different promoter specificities. Dimerization is inhibited by the hsp90 heat shock protein.	29, 247, 248
PRDI-BF1h		AAGTGAAAGT	88 <sup>c</sup> zinc f.		virus infection	Potent repressor of INFB gene expression.	249
PTF		TCAGAGTATAAATACT	40 <sup>b</sup>	pituitary		May be a modified form of TFIID. Responsible for pituitary-specific expression of the transcription factor GHF1.	250
PTF1m,r		ATGGGAN <sub>4</sub> CTCAGCTGTGC	64 <sup>a</sup> + 48 <sup>a</sup>	fetal and adult pancreas (from day 15 of gestation)		May be implicated in the coordinate expression of genes transcribed in the acinar pancreas and in differentiation of the pancreas.	251, 252, 253
Pu.1ch	Sfp1-1ch	AGAGGAACT		Most abundant in B lymphocytes and macrophages		Related to members of the Ets family. Might block preerythroblast differentiation, thereby causing immortalization.	254, 255
Pu1h		CGGTGG				Acts on c-myc gene promoter.	256





Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa): <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
TBP <sub>h,m,r</sub>		TATAAA	h: 38 <sup>a</sup> m: 35 <sup>a</sup>	ubiquitous		TATA box-binding factor. TBP is part of the TFIID protein complex implicated in RNA polymerase II positioning. TFIID-binding to the TATA box is stimulated by TFIIA and TFIIB. Binds as a monomer.	281, 282, 283, 284
TCF-1 <sub>h</sub>	TCF-1 <sub>h</sub> LEF-1 <sub>m</sub>	C <sub>1</sub> AA <sup>c</sup> C <sub>1</sub> AAG	53 <sup>a</sup> -57 <sup>a</sup> HMG	T cells		Distally related to the Ets family. Exists in three alternative splice forms. The DNA-binding domain is similar to that of high mobility group proteins.	275, 276, 277, 278
TCF-2 <sub>h</sub>		G <sub>1</sub> CAGGAAGT <sup>c</sup> /C	63 <sup>a</sup>	T cells		Member of the Ets family.	277
TEF-1 <sub>h</sub>		AAGT <sup>c</sup> /CATGCA TGGAAATGT	53 <sup>a</sup>			Binds cooperatively to the Sph- and tandem repeats of the GTTIC motifs on the SV40 promoter.	279, 280
TEF-2 <sub>h</sub>		GGGTGTGG	57 <sup>a</sup>			Binds to the GT-1C motif on the SV40 promoter. May be related to AP3.	279, 24
TFE3 <sub>m</sub>		GCCACATGACC	59 <sup>c</sup> b/HLH/ZIP			Binds to IgH $\mu$ E3 motif and the USF binding site. Probably different from NF- $\mu$ E3. Closely related to TFEB.	73, 200
TFEB <sub>h</sub>		GGCCACGTGACC	60 <sup>a</sup> b/HLH/ZIP			Closely related to TFE3. Binds to E boxes of IgH gene promoters.	285
TGT3 <sub>h</sub>		AAGTGTTC		liver		Acts on hepatitis B virus enhancer.	286
TIN-1 <sup>r</sup>		AGGAAGTTCC	43 <sup>a</sup> /45 <sup>a</sup>	testis		Acts as a transcriptional inhibitor.	287

Factor 1	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa); <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
TRh.s,m,ch	c-ErbA <sup>h,m,r,ch</sup>	(AGGTCA) <sub>2</sub>	46 <sup>a</sup> zinc f. (Ph)	ubiquitous	thyroid hormone	Thyroid hormone receptor. Member of the steroid hormone receptor superfamily. Vertebrates express two different forms, TRα and TRβ, with different functions in early embryonic brain as well as during the late phase of hormone requirement. Binds to DNA as a monomer or as a homodimer. Binds to head-to-head or tail-to-tail palindromes or to direct repeats of the half-site shown.	29, 288, 289, 290, 291, 292
TTF-1 <sup>h,m,r,c</sup>	TgTF-1 <sup>r</sup> T/EBP <sup>r</sup>	GNNCACTCAAG	38 <sup>a</sup> HD	thyroid, lung	(ras)	Is largely responsible for thyroid-specific gene expression and may be implicated in the differentiation of the thyroid.	293, 294, 295
UBP-1 <sup>h</sup>		CTCTCTGG	61 <sup>a,63a</sup>			Binds to the HIV tar element, but also to the TATA box.	86
USF <sup>h,m,r</sup>	MLTF <sup>h,m,r</sup> B1 <sup>f</sup>	GGG/CCACG/ATGAC	43 <sup>a,44a</sup> b/HLH/ZIP	ubiquitous	high mobility group proteins 1 and 2 (DNA methylation)	Heat-stable, may be implicated in the regulation of tissue-specific and developmentally regulated genes. Has been suggested to be related to the centromere binding protein CBF-1. USF mRNA is alternatively spliced. Increases rate or stability of TFIID binding during <i>in vitro</i> chromatin assembly.	50, 296, 297, 298, 299, 300
VBp <sup>ch</sup>		GTTTACATAAAC	bZIP	ubiquitous, most abundant in the oviduct		Appears to play a pivotal role in the estrogen-dependent regulation of the chicken vitellogenin gene.	301
VDR <sup>h,m,r,ch</sup>		AGGTCATGACCT	h: 48.3 <sup>c</sup> ch: 58 <sup>a,60a</sup> zinc f. (Ph)		vitamin D	Vitamin D receptor, member of the steroid hormone receptor superfamily.	242, 302

Factor <sup>1</sup>	Synonyms, Homologs <sup>2</sup>	Binding sequence <sup>3</sup>	M <sub>r</sub> (kDa), <sup>4</sup> domains <sup>5,6</sup>	Tissue specificity	Inducers (repressors) <sup>7</sup>	Features	References
WT-ZFph	WT1m	CGCCCCCGC	h: 21.5a m: 47/49c zinc f.	developing kidney and Bowman's capsule, meta- nephric blastema		Encoded by the Wilms' tumor locus. Could act as antagonistic to EGR-1 and EGR-2. Reveals homology to EGR-1. Exists in two alternative splice forms. May regulate transcription during nephroblast development. Acts as a transcriptional repressor.	303, 304, 305, 306
XFI/2m,r		TCTTCTCAGCAACT			(cycloheximide)	Binds to the xenobiotic response element (XRE) of the cytochrome P450c (CYP1A1) gene promoter. Not responsive to polycyclic compounds. Distinct from Ahr.	307
XPF-1m		CACCTGN <sub>4</sub> TTTCCC	60a	exocrine pancreas		Implicated in transcriptional activation of exocrine pancreas-specific gene expression.	308
YB-1h		ATTTTCTGATTGGCCAAAG	35c		(PMA, INFγ)	May be a Y-box binding protein, but is also suspected to be involved in maintenance of chromatin structure or DNA repair as a non-specific DNA- binding protein.	309, 315

## Nucleotide Sequence, Transcriptional Analysis, and Glucose Regulation of the Phenoxazinone Synthase Gene (*phsA*) from *Streptomyces antibioticus*

CHUIN-JU HSIEH AND GEORGE H. JONES\*

Department of Biology, Emory University, Atlanta, Georgia 30322

Received 27 June 1995/Accepted 8 August 1995

The nucleotide sequence of a 2.3-kb *SphI* fragment containing the structural gene (*phsA*) for phenoxazinone synthase (PHS) of *Streptomyces antibioticus* was determined. The sequence was found to contain an open reading frame (ORF) with a G+C content of 71.5% oriented in the direction of transcription that was confirmed by primer extension. The ORF encodes a protein with an  $M_r$  of 70,223 consisting of 642 amino acids and is preceded by a potential ribosome-binding site. The codon usage pattern is in agreement with the general pattern for streptomycete genes, with a 92.5 mol% G+C content in the third position. The N-terminal sequence of the mature PHS subunit corresponds exactly to that predicted from the nucleotide sequence. Neither ATG nor GTG initiator codons were identified for the protein. However, a TTG codon was located near the amino terminus of the mature protein and is a good candidate for the initiator codon. The transcriptional start point of *phsA* was located 36 bp upstream of the start codon by primer extension. The  $-10$  region of the putative promoter showed some similarity to the consensus sequence for the major class of prokaryotic promoters, but the  $-35$  region was less similar. Comparison of the primary amino acid sequence of PHS of *S. antibioticus* with other amino acid sequences indicated that PHS is a blue copper protein with copper binding domains in the N-terminal and C-terminal regions of the polypeptide chain. A *BsrBI* fragment containing the promoter region of *phsA* and a portion of the ORF was shown to promote *xylE* expression when cloned in the streptomycete promoter probe vector pIJ2843. This *phsA* promoter-dependent *xylE* expression could be repressed by glucose in *S. antibioticus* when the organism was grown on glucose or galactose plus glucose. Thus, the cloned promoter region appears to contain the sequences responsible for catabolite repression of PHS production.

Actinomycin is one of the antibiotics produced by the gram-positive actinomycete *Streptomyces antibioticus* (52). A putative pathway for actinomycin biosynthesis was proposed several years ago (50), and biochemical, physiological, and genetic studies have confirmed the essential details of that pathway. Five enzymes from *S. antibioticus*, *Streptomyces chrysomallus*, and *Streptomyces parvulus* have been isolated and characterized to demonstrate their involvement in the actinomycin biosynthetic pathway (6, 11, 22, 23, 28–31). One of these enzymes, phenoxazinone synthase (PHS), catalyzes the oxidative condensation of two molecules of 4-methyl 3-hydroxyanthraniloyl pentapeptide to form actinomycinic acid, which is the penultimate intermediate in the putative biosynthetic pathway (Fig. 1). The enzyme was first identified by Katz and Weissbach (28) and subsequently purified by Choy and Jones (6). To date, *phsA*, the gene coding for PHS from *S. antibioticus*, is the only gene involved in actinomycin biosynthesis that has been cloned (25).

Although essentially all of the enzymes required for actinomycin production have been identified, little is known about the regulation of these enzymes and of overall actinomycin production. Of all the enzymes identified, PHS is perhaps the best characterized. It was shown some years ago by Marshall and coworkers that actinomycin production is repressed in *S. antibioticus* cultures grown on glucose or galactose plus glucose as compared with cultures grown on production medium with galactose alone as the carbon source (39).

Catabolite control has been implicated in the expression of both PHS and actinomycin synthetase I (ACMSI; the enzyme

which activates the precursors of the actinomycin chromophore in *S. antibioticus* [23, 30, 31]). PHS production was demonstrated to be subject to catabolite control shortly after the identification of the enzyme (13, 28). It is possible that *phsA* and *acmsI* are located in the same genomic region in *S. antibioticus* since it was well known that the genes for antibiotic production are clustered in the streptomycete genome (for examples, see reference 38). Therefore, the detailed molecular analyses of the mechanisms controlling the expression of *phsA* are essential to our understanding of the regulation of actinomycin biosynthesis and the synthesis of other antibiotics. We report here the nucleotide sequence and transcriptional analysis of *phsA* and identify the promoter region of the gene. We also demonstrate that a cloned fragment containing the putative promoter is active in a streptomycete promoter probe vector and that the activity of the promoter is repressed when *S. antibioticus* transformants containing the relevant constructs are grown on glucose or galactose plus glucose as compared with cultures grown on galactose as the sole carbon source.

### MATERIALS AND METHODS

**Organisms and growth conditions.** The *Streptomyces* strains used were *S. antibioticus* IMRU 3720 and *Streptomyces lividans* 66 derivative TK24 (18). *S. antibioticus* was grown on liquid NZ-amine and galactose-glutamic acid media as described previously (13). *S. lividans* was generally grown on yeast extract-malt extract plus 34% sucrose (YEME) or on tryptone soy broth. For protoplast preparation, TK24 was grown on YEME with  $MgCl_2$  and glycine at the final concentrations of 5 mM and 0.5%, respectively (49). Protoplasts were allowed to regenerate on R2YE medium (49) for 12 to 24 h and then overlaid with 2 to 3 ml of soft nutrient agar supplemented with thiostrepton at a final concentration of 500  $\mu$ g/ml. Tyrosine at 0.075% (wt/vol) was added to the soft nutrient agar for overlaying when pIJ702 derivatives were used.

*Escherichia coli* DH5 $\alpha$  [ $F^-$   $\phi$ 80 *dlacZ*M15 (*lacZ*Y $\alpha$ -arg $f'$ ) U169 *endA1* *recA1* *hsdR17* ( $r_K^+$   $m_K^+$ ) *deoR* *thi-1* *supE*441 *gyrA96* *relA1*] and XL-1 Blue 2 [*recA1*

\* Corresponding author. Phone: (404) 727-4208. Fax: (404) 727-2880. Electronic mail address: GJONES@BIOLOGY.EMORY.EDU.

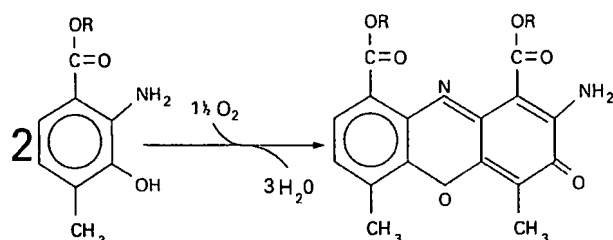


FIG. 1. The PHS reaction. The penultimate step in the actinomycin biosynthetic pathway in *S. antibioticus*, the oxidative condensation of two molecules of 4-methyl-3-hydroxyanthranilic acid to yield actinomycin acid, is catalyzed by PHS.

*endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* (F' *proAB lacPZ M15 Tn10* (Tet')) were generally cultured in L broth or on L agar (35). *E. coli*-competent cells were prepared by the  $\text{CaCl}_2$  method and transformed as described by Sambrook et al. (46). After transformation with pUC19 and pBluescript SK<sup>+</sup> derivatives, transformants were selected on L agar plates containing 100  $\mu\text{g}$  of ampicillin per ml, 40 mg of 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal) per ml, and 0.2 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). For single-stranded DNA preparation, strains containing pBluescript SK<sup>+</sup> derivatives were grown in 2XYT medium in the presence of 100  $\mu\text{g}$  of ampicillin per ml and helper phage VCSM13 for 2 h followed by the addition of 75  $\mu\text{g}$  of kanamycin per ml and growth overnight. Growth temperatures for *Streptomyces* spp. and *E. coli* were 30 and 37°C, respectively.

**DNA manipulations.** Plasmid and chromosomal DNAs were prepared as described previously (4, 17, 20) and analyzed by restriction digestion and agarose gel electrophoresis. In some experiments, restriction fragments were recovered from low-melting-point agarose as described by Favre (10). Protoplast preparation, transformation, and regeneration were as described previously (17, 20, 25). A list of plasmids used or generated in the present study is provided in Table 1. pJSE923 is a derivative of pIJ2501 (25) with an *Xba*I linker inserted at the *Pvu*I site of *phsA*. pJSE929 contains the blunt-ended *Bsr*BI subfragment of the *phsA* promoter region cloned into the *Hinc*II site of pUC19. pJSE935 contains the *Hind*III-*Bam*HI subfragment of the *phsA* promoter region of pJSE929 cloned into *Hind*III-*Bam*HI-digested pIJ2843 (7).

**Enzyme assays.** *Streptomyces* cultures were grown in 250-ml flasks containing 50 ml of glutamic acid-salts medium, 50  $\mu\text{g}$  of thiostrepton per ml as necessary, and 5 mM  $\text{CuSO}_4$  at 28°C with shaking at 200 rpm. Cultures contained either 1% galactose, 1% glucose, or 0.5% galactose plus 0.5% glucose as carbon sources. The cultures were harvested 12 h after inoculation. Mycelium was washed in 100 mM potassium phosphate (pH 7.5), suspended in a final volume of 2 ml of sample buffer (19), and disrupted by sonication.

Catechol dioxygenase assays were performed and activities were determined spectrophotometrically as described previously (19, 54). Catechol dioxygenase specific activity was calculated as the rate of change in  $A_{375}$  per min per milligram of protein and converted to millunits per milligram (45). Protein concentrations were determined with the bicinchoninic acid protein assay reagent kit from Pierce. The PHS assay was performed as described previously (6) with 3-hydroxyanthranilic acid as the substrate.

**Nucleotide sequence analysis.** Sequential deletion clones from both ends of the *phsA* *Sph*I fragment were obtained by exonuclease III-mung bean nuclease digestion with the exonuclease III-mung bean deletion kit from Stratagene Clon-

ing Systems. The *phsA* *Sph*I fragment was subcloned into pBluescript SK<sup>+</sup> (Stratagene) modified to contain an *Sph*I site in the polylinker, and the resulting recombinant plasmids (pJSE900 and pJSE910) were used to create deletion clones suitable for sequencing. The nucleotide sequences of both DNA strands of the cloned *phsA* fragment were obtained by the dideoxy chain termination method (47). Single-stranded DNA was obtained with VCSM13 as a helper phage, and the DNA was prepared as described previously (26). The sequencing reactions were performed basically as described for the 7-deaza-GTP Sequenase kit from United States Biochemicals except that the extension and termination reactions were done at 50 and 70°C, respectively. The reactions were post-terminated at 70°C for 2.5 min by adding 2.5 U of *Taq* version 2.0 DNA polymerase and 1  $\mu\text{l}$  of termination mixture, both from United States Biochemicals. Difficult compression areas and pause sites were resolved by using dITP instead of deaza-GTP. The DNA sequences were analyzed with the DNAsis program from Hitachi and the GCG program from the University of Wisconsin.

The GenBank accession number for the *S. antibioticus* IMRU3720 PHS gene (*phsA*) is U04283.

**Primer extension.** In the primer extension experiments, a 24-base oligonucleotide primer, 5'-GATCTCGGTCTCCCGCTCACCTC-3', that is located 528 bp downstream of the 5'-*Sph*I site and is complementary to the *phsA* mRNA was used to reveal the transcriptional start point. End labeling of the 5'-terminus of the oligonucleotide primer with the polynucleotide kinase reaction and the primer extension reaction were done as described by Moran (42). RNA preparation was as described previously (17) with the following modifications. Mycelium was collected on a Whatman no. 4 filter disc by use of a vacuum line to accelerate the filtration process. The mycelium was quickly scraped off the filter into a universal bottle and resuspended in 5 ml of modified Kirby mixture at 4°C (modified Kirby mixture consists of 1% [wt/vol] sodium trisopropylphenylsulfonate [Eastman Chemicals], 6% [wt/vol] sodium-4-amino salicylic acid [sodium salt; BDH], and 6% [vol/vol] Tris-EDTA-buffered phenol mixture, and all solutions were made up in 50 mM Tris-HCl [pH 8.3]). The contents were vortexed with 10 g of 4.5- to 5.5-mm-diameter glass balls as vigorously as possible for at least 2 min. Three milliliters of phenol-chloroform mixture was added, and the mixture was vortexed as described above. The homogenate was then transferred to a polypropylene tube (Falcon 2006) and centrifuged (10 min at 12,000  $\times$  g and 4°C) to separate the phases. The aqueous layer was transferred to a fresh tube, and an additional 5 ml of phenol-chloroform mixture was added. The solutions were vortexed thoroughly for 2 min and centrifuged again as described previously to separate the phases, and this procedure was repeated until very little interphase material remained visible. One-tenth volume of 4 M sodium acetate (pH 6.0), followed by an equal volume of isopropanol, was added to the aqueous phase. The solutions were mixed and left at -20°C for 1 h. The nucleic acids were collected by centrifugation at 12,000  $\times$  g for 10 min, and the supernatant was discarded. The pellet was rinsed with absolute ethanol and vacuum dried. The pellet was resuspended in 180  $\mu\text{l}$  of distilled water (treated with diethyl pyrocarbonate) and 20  $\mu\text{l}$  of 10 $\times$  DNase buffer (0.5 M Tris-HCl [pH 7.8], 0.05 M  $\text{MgCl}_2$ ) and transferred to an Eppendorf tube. DNase (RNase-free; Sigma Chemical Co.) was added to a final concentration of 30  $\mu\text{g}/\text{ml}$ . The solutions were incubated at room temperature for 30 min. An equal volume of phenol-chloroform mixture was then added, and the samples were mixed by vortexing. The phases were separated by centrifugation in a microfuge, and the aqueous phase was transferred to a fresh tube. The aqueous phases were then extracted by adding an equal volume of chloroform. Total RNA was precipitated with 1/10 volume of 3 M sodium acetate (pH 6) and an equal volume of isopropanol for 2 h at -20°C, and the precipitate was collected by centrifugation. The RNA pellet was rinsed with 70% and then 100% ethanol, vacuum dried, resuspended in 100  $\mu\text{l}$  of distilled water, and stored at -70°C. The quantity of RNA was assessed by spectrophotometry, and the quality was assessed by agarose gel electrophoresis.

TABLE 1. Plasmids used or referred to in the present study

Plasmid	Description	Source or reference
pUC19		53
pBluescript SK <sup>+</sup>	Phagemid cloning vector (Stratagene); the vector was modified to contain an <i>Sph</i> I site in the polylinker	
pIJ702		27
pIJ2501	The 2.3-kb <i>phsA</i> <i>Sph</i> I structural gene from <i>S. antibioticus</i> cloned into the <i>Sph</i> I site of pIJ702	25
pIJ2843	<i>Streptomyces</i> low-copy-number promoter-probe vector	7
pJSE900 and pJSE910	The 2.3-kb <i>phsA</i> <i>Sph</i> I cloned in the <i>Sph</i> I site of pBluescript SK <sup>+</sup> in two orientations	Present study
pJSE923	pIJ2501 with an <i>Xba</i> I linker at the <i>Pvu</i> I site of <i>phsA</i>	Present study
pJSE929	The blunt-ended, ca. 235-bp <i>Bsr</i> BI subfragment of the <i>phsA</i> promoter region, extending from position -106 to +135 relative to the transcriptional start site, cloned into the <i>Hinc</i> II site of pUC19	Present study
pJSE935	The ca. 265-bp <i>Hind</i> III- <i>Bam</i> HI subfragment of the <i>phsA</i> promoter region of pJSE929 cloned into <i>Hind</i> III- <i>Bam</i> HI-digested pIJ2843	Present study

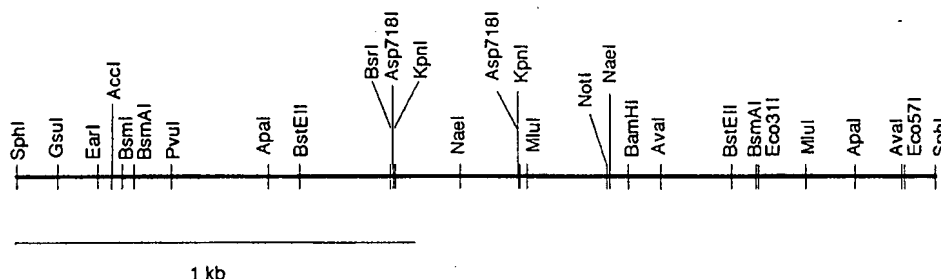


FIG. 2. Restriction map of *phsA* constructed from the *phsA* sequence.

**Determination of the amino-terminal sequence of the PHS subunit.** The amino-terminal sequences of the cloned and native PHS proteins were determined at the Emory University Microchemical Facility and found to be identical. The sequence of the first 15 amino acids of the protein is Thr-Asp-Met-Ile-Glu-Gln-Ser-Asp-Asp-Arg-Ile-Asp-Pro-Ile-Asp.

**Enzymes and reagents.** Restriction endonucleases were purchased from Boehringer-Mannheim Corporation, Gibco BRL, and Promega Corporation. Calf intestinal alkaline phosphatase, T4 DNA ligase, and avian reverse transcriptase were obtained from United States Biochemicals. Exonuclease III and mung bean nuclease were obtained from Stratagene. Sigma Chemical Co. supplied RNase, which was prepared as described previously (46). The 7-deaza-dGTP Sequenase version 2.0 and *Taq* version 2.0 DNA polymerase kits were purchased from United States Biochemicals. [ $\gamma$ - $^{32}$ P]dATP, [ $\alpha$ - $^{32}$ P]dCTP, and  $\alpha$ - $^{35}$ S-dATP were purchased from Dupont New England Nuclear Products and Amersham. RNasin was obtained from Promega. All of the chemicals were of reagent grade or the highest purity commercially available.

## RESULTS

**Nucleotide sequence analysis.** A detailed restriction map of the *phsA* *SphI* fragment constructed on the basis of the nucleotide sequence is shown in Fig. 2, and the nucleotide sequence of the fragment is shown in Fig. 3. Analysis of the DNA sequence with the FRAME codon preference program (3) revealed a 1,932-bp open reading frame with 71.5% G+C content, matching the codon usage of *Streptomyces* spp. (Fig. 4). The open reading frame presumably starts with a TTG codon at nucleotide 348 and encodes a deduced polypeptide of 642 amino acids with a predicted  $M_r$  of 70,223. Furthermore, the predicted initiator amino acid is only one position upstream of the N-terminal amino acid obtained by protein sequence analysis of purified PHS (the first 15 amino acids shown in Fig. 3; see Materials and Methods). Additional information on the putative translational start was obtained by inserting an *XbaI* linker downstream of this region (Table 1, pJSE923). The inserted linker created stop codons in all three reading frames. When the resulting recombinant plasmid, containing the *XbaI* linker in the *PvuI* site of *phsA*, was used to transform *S. lividans*, PHS expression from *phsA* was completely abolished (data not shown). These results rule out the possibility that the cloned fragment activates a normally silent *phsA* gene in *S. antibioticus*, as has been observed for *S. lividans* (25, 37). Upstream of the putative TTG start codon is the sequence GGGGG (Fig. 3, boxed), which may act as a ribosome binding site (48). A short stem-loop structure is located 4 bp downstream of the *phsA* stop codon (Fig. 3, inverted arrows), but its ability to function in transcription termination is problematic because of its length.

**Primer extension analysis and identification of the putative *phsA* promoter.** A 24-mer oligonucleotide primer, corresponding to sequences 530 bp downstream of the 5' *SphI* site and 180 bp downstream of the translational start codon (Fig. 3), was used in primer extension studies to locate the 5' end of the *phsA* transcript (Fig. 5). RNA templates were prepared from *S. antibioticus* and *S. lividans* as indicated in the legend to Fig. 5.

The transcriptional start point (*tsp*) of the *phs* message revealed by this analysis is located at the A residue which is 313 bp downstream of the 5' *SphI* site and 36 bp 5' to the translation initiation codon. The transcription start point of the cloned *phsA* gene in *S. lividans* TK24 is the same as that of the chromosomal gene in *S. antibioticus* (Fig. 5). In addition, there is no difference in the *tsp* shown in the primer extension experiments using total RNA prepared from glucose- or galactose-grown cultures (data not shown). However, glucose-grown cultures contained less *phs*-specific message than galactose-grown cultures. This observation is consistent with earlier data suggesting that the decreased level of PHS observed in cultures grown on glucose as compared with that in galactose-grown cultures is due in part to an effect at the level of *phs* transcription (20, 21).

On the basis of primer extension studies, putative -10 and -35 promoter regions were located relative to the transcription start point (Fig. 3). There are also other interesting features which are located near the promoter region, including several sets of direct repeat sequences, two sets of inverted repeats, and two TNTNAN sequences (Fig. 3). These sequences are noteworthy because they may be involved in the catabolite control of the *phsA* gene (41). The function of these sequences will be examined in detail in subsequent studies.

Confirmation of the presence of a functional promoter upstream of the transcription start site was obtained by promoter probe cloning. In these experiments, a *BsrBI* fragment from *phsA* (see Fig. 2 and 3) was inserted upstream of the *xyIE* gene in the promoter probe vector pIJ2843 (7, 36). The resulting recombinant plasmid was used to transform *S. antibioticus* and *S. lividans*, and mycelial extracts were prepared after 19 h of growth of control and transformed cultures in liquid media. The results of catechol dioxygenase assays of those extracts revealed that the untransformed strains contained negligible levels of enzyme activity, as was also the case for strains transformed with pIJ2843. In contrast, *S. antibioticus* and *S. lividans* strains containing pJSE935, with the putative promoter fragment, showed significant levels of *xyIE* activity (data not shown). Thus, the *BsrBI* fragment does possess promoter activity, and the promoter probe results support the identification of the promoter region of *phsA* suggested by the sequencing and primer extension studies. The use of pJSE935 in studies of glucose repression of *phsA* is described below.

**Sequence comparisons with PHS sequence.** The deduced amino acid sequence of PHS was compared with entries in protein databases provided by GenBank by use of the FASTA program. The sequence with the greatest homology to PHS was that of bilirubin oxidase from *Myrothecium verrucaria* (32). There is 26% identity and 45% similarity between the sequences of PHS and the bilirubin oxidase protein. A lower homology (18% identity, 40% similarity) was found for the

GGCATGCAGAGTTCGAGACATCATGCCCAAGAACTCCCGCTCTGTACGACGACTTC 60  
 GACCTOTCGGCGAGGCGCGCGCGGATGACAACTACGTGGCGGACCTCCAGAACTTC 120  
 GTGTGGGATCTCTCAATGGGATCGCGGGTGGACGTTACAAGTCCGCGTACCTGGCG 180  
 AGCCGGGCGCAGGGTTCGTCCGAGCGCTCTTCGGCCACTCCGCCCCACACCTGCTG 240  
 CTTTCGGGACTTTTCATATACGGAATAGGGCATTCAGGGAACCGCGAGGGGCGCGGGCG 300  
 TCTCATCGGGCAAGGGAACGACCGACGAACACGGGGGCTTCGACCTTACCGACATG 360  
 ATCGAGCAGAGCGACGACCGATCGAGCGGATCGCGGAGTCTTGGCGAGCGGACTCTTG 420  
 I E O S D D R I D P I D G V L A D G V L  
 GCGGACGAGCTTCTGGGAGGAGCGGAGGACCGCGCGCGGTAACGACACCG 480  
 TACCGGCGGAGTTCGACCTCCGCGCGCTCTCGCGCGCGCTTCGAGCGAGGTGACCGG 540  
 GAGACCGAGATCGCGCTGCGCGCGCGCTCGCGCGCGCTCGCGCGCGCTCGCGCGCG 600  
 C T E I A L R P T W V R L H P Q L P P T  
 CTGATCTGGGCTACGACCGGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
 L M W G Y D G Q V P P G P T I E V R R C Q  
 CGCGTCCGATCGCGTGAACCAACCGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
 R V R I A W L P P T W V R L H P Q L P P T  
 GTGAGGTGCGCGCTCG 780  
 V E V P L G N T P P T P A P N T E P G R G  
 GCGTGAACCAACGAGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
 G V E P N K D V A A L P A W S V T H L H  
 GCGCGGACGAGCGCGGGAACGAGCGGCTGGCGGAGAACGCGCTCGGCTTCGGCGAC 900  
 G A Q T G G G N D G W A D N A V G F G D  
 Domain 1  
 GCCAGCTCTCGAGTATCGGAACGACGAGCGACCGAGCGAGTGTGTACGACGAC 960  
 A Q L S E Y P N D H Q A T Q W W Y H D H  
 Domain 2  
 GCCATGAACATCACCGGTGGAACGTGATGGCGGCGCTGTACCGCACTACCTGCTCG 1020  
 A M N I T R N W N V M A G L Y G T Y L V R  
 GACGAGGAGGAGCGCGCTCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080  
 D D E E D A I G L P S G D R E I P L L I  
 GCGGCGGCAACCTCGACACCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140  
 A D R N L D T D E D G R L N G R L L H K  
 ACGGTGATCGTCCAGCGTTCACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200  
 T V I V Q Q S N P E T G K P V S I P P F  
 GCGCGTACACCGCGTCAACCGCGGATCTGGCGGATCGCGCGCGCGCGCGCGCGCG 1260  
 G P Y T T V N G R I W P Y A D V D D G W  
 TACCGGCTCGCGCTGTGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320  
 Y R L R L V N H A S N A R I Y N L V L I D  
 GAGGACGAGCG 1380  
 E D D R P V G V W H Q I G S D G G L L  
 CCGCGCGCGTCCGCTGACTGACGACGACCTCGCGCGCGCGCGCGCGCGCGCGCG 1440  
 P R P V P D F D D T L P V L S A A P A  
 GAGCGGTTCGACCTGTGTGCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
 E R F D L L V D F R A L G G R R L R L V  
 BamHI  
 GACAAGGGGCG 1560  
 D K G P G A P A G T P D P L G G V R Y P  
 GAGGTGAGAGTTCGCGGTTCGCGGAGACTTCGCGGAGGAGTCTTCGCGCTCCGCG 1620  
 E V M E F R V R E T C E E D S F A L P E  
 GTGCTCTCGCGTCTTCGCGCGGATGAGCGACGACATCCGCGCGCGCGCGCGCGCG 1680  
 V L S D I F R W S H D J P H G H E L I  
 GTGCTGACCG 1740  
 V L T P P G T K G S G H F E I W E M A  
 GAGGTGAGGACCG 1800  
 E V E D P A D V V P A E G V I Q V T G  
 GCGGCGCGCGTCAAGACGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860  
 A D G R T K T Y R R T A A T P N D G L G  
 TTCACATCGCGGCG 1920  
 P T I G E G T H E Q W T F L N L S P I L  
 CACCCATGACATTCACCTGGCGGACTCCAGGTCTCTCGCGCGCGCGCGCGCGCGCG 1980  
 H P M H I H L A D P Q V L G R D A Y D A  
 Domain 3  
 TCGCGCTTCGACCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040  
 S G F D L A L G G T R T P V R L D P D T  
 CCGGTCCGCTCG 2100  
 P V P L A P N E L G H K D V F Q V P G P  
 CAAGGGCTGCGGTGATGGGCAAGTTCGAGCGGCGGTACGGCGGTTCATGTACCACTG 2160  
 Q G L R V M G K F D G A Y G R F M Y H C  
 CACCTCTCGAAGCAGGAGATGGGATGATCGCGCGCGCGCGCGCGCGCGCGCGCG 2200  
 H L L E H E D H M M R P F V M M P P E  
 Domain 4  
 GCCTGAAGTTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2280  
 A L K F D H G G A H G G H G E G H T G \*  
 CGCCCG 2302

FIG. 3. Nucleotide sequence of the *SphI* fragment containing the *S. antibioticus phsA* gene. Important restriction enzyme sites are indicated above the sequence. The boxed region denotes the possible ribosomal binding site (Shine-Dalgarno sequence [SD]). The potential stem-loop is indicated by a pair of inverted arrows located toward the end of the sequence. Potential -10 and -35 regions have lines above the sequence and are discussed in the text. The location of the transcription start point is indicated by an upward arrow and bold type, and the primer for primer extension is shown by the long arrow under the nucleotide sequence around position 530. The direct and inverted repeat sequences are indicated by arrows and numbered in pairs (1 to 8). The TNTAN elements are shown in bold type and are located around 253 and 303 bp from the 5' *SphI* site. Four potential copper binding domains are also indicated in the sequence (solid bars). These sequence data appear in the EMBL, GenBank, and DDBJ nucleotide sequence data libraries under accession number U04283.

sequence of manganese-oxidizing protein from *Leptothrix discophora* (8). Copper binding motifs of all three proteins are aligned in Fig. 6. All three proteins are involved in oxidation reactions, but only PHS and bilirubin oxidase belong to the family of blue copper proteins (2, 12, 32). Sequence comparison of PHS with bilirubin oxidase, manganese-oxidizing protein, and several other blue copper proteins revealed the presence of four regions in the sequence of the former protein corresponding to the potential copper binding domains found in the sequences of the blue copper proteins (Fig. 6). The finding of these copper binding domains confirms PHS as a blue copper protein (2, 12). This result is not at all surprising, since PHS has been shown to require copper for activity (2). The amino acid sequence of PHS contains consensus domains for the copper binding regions of the same types (I, II, and III), which were revealed by X-ray crystallography of ascorbate oxidase from zucchini (40). However, there are just two copper binding domains found in the manganese-oxidizing protein. We speculate that the copper binding domains are components of the catalytic sites of these enzymes.

Expression of the cloned *phsA* promoter is repressed by glucose in *S. antibioticus*. The production of PHS in *S. antibioticus* was demonstrated some years ago to be subject to catabolite control (13, 28). As has been mentioned, later studies suggested that the production of PHS was regulated at the transcriptional level (20, 21). In the present study, the effects of glucose on the expression of the promoter active fragment cloned in pJSE935 was examined in *S. antibioticus*. Transformants containing pJSE935 were grown on 1% galactose, 1% glucose, or a mixture of 0.5% galactose and 0.5% glucose. Catechol dioxygenase assays were performed on extracts of mycelium harvested 12 h after inoculation of the growth media. PHS assays were performed on these same extracts. The results of these experiments, presented in Fig. 7, show that glucose represses the expression of the *phsA* promoter in pJSE935 in the presence or absence of galactose. Thus, the effects of glucose on the *phsA* promoter would seem to fit the classical definition of catabolite repression, which requires that expression of the relevant gene be inhibited when the organism in question is grown on the repressing and (relatively) nonrepressing carbon sources simultaneously. It is significant that the PHS activity in the mycelial extracts exactly paralleled the *xylE* activity; PHS production was inhibited when the organism was grown on glucose alone or on galactose plus glucose (Fig. 7).

One possible mechanism for catabolite repression of *phsA* expression would involve the binding of a repressor protein to operator sequences in the promoter region of the gene. Such mechanisms have been suggested to explain glucose repression in other streptomycetes (for examples, see references 9 and 51). To examine this possibility in *S. antibioticus*, the effects of carbon source on PHS activity were measured in transformants containing pJSE923, in which the *phsA* gene is disrupted by an *XbaI* linker. As controls, PHS activity was measured in untransformed *S. antibioticus* and in transformants containing pIJ702 and pIJ2501. We reasoned that if the *phsA* promoter region possesses a repressor binding site, it might be possible to titrate the repressor by cloning that site at high copy in *S. antibioticus*. However, the result of the experiment was the observation that transformants containing pJSE923 showed the same pattern of PHS expression when grown on galactose, glucose, or glucose plus galactose as did the wild-type strain (data not shown). Thus, although the *XbaI* linker effectively prevented expression of the cloned *phs* gene, the presence of the disrupted gene at high copy did not abolish glucose repression of the endogenous *phsA* gene.



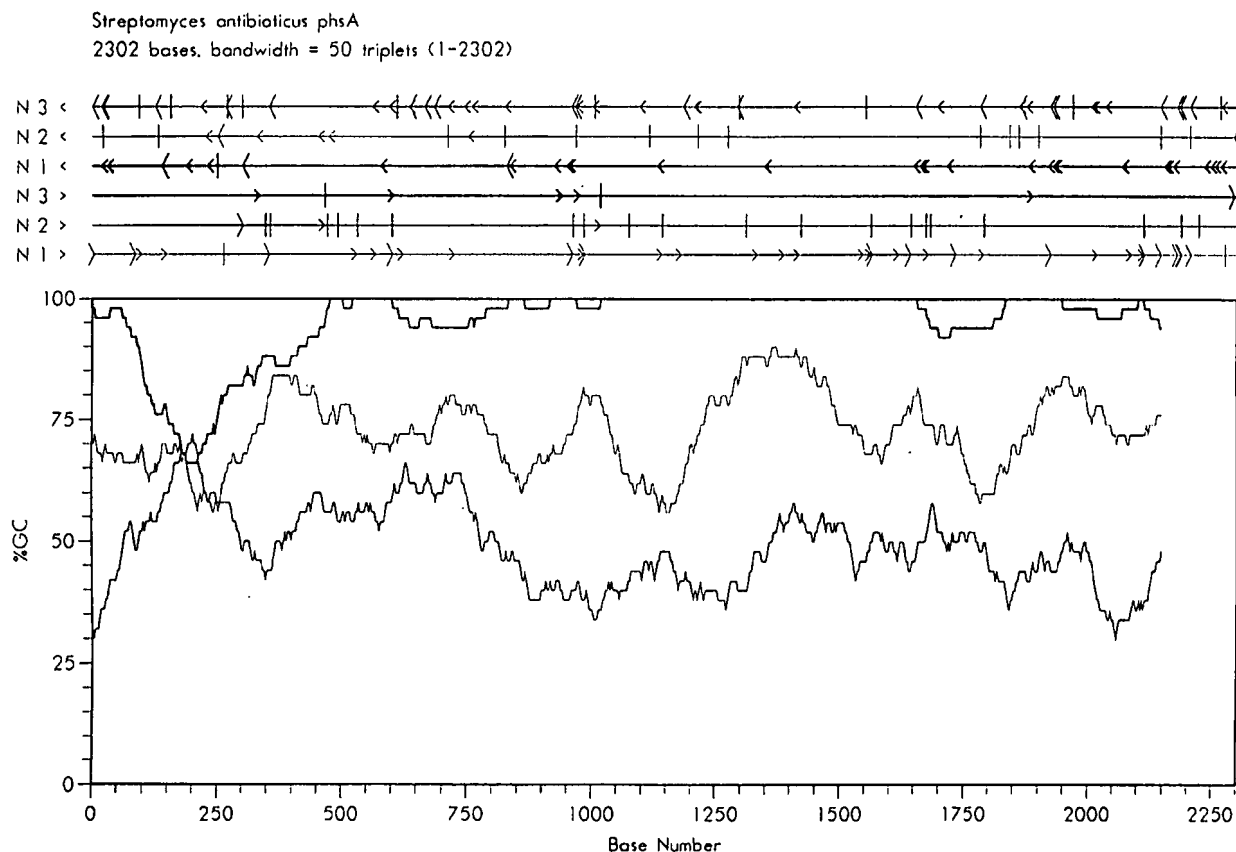


FIG. 4. Analysis of the DNA sequence with the FRAME program (3) revealed a 1,932-bp open reading frame matching the codon usage of *Streptomyces* spp.

## DISCUSSION

In the present study, we have characterized the cloned PHS gene from *S. antibioticus*. The  $M_r$  of the PHS subunit deduced from the nucleotide sequence data is 70,223. This value differs from the apparent value of 88,000 estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (25). The explanation for the anomalous migration of this protein on SDS-PAGE is not clear. Previous studies have not revealed the presence of carbohydrate or other substances covalently associated with PHS, but other features of the protein presumably cause it to migrate in an unexpected fashion. Two native forms of PHS, large and small (L and S), were reported previously to have  $M_r$ s of 540,000 and 180,000 and to be composed of six and two PHS subunits, respectively (6). On the basis of the deduced  $M_r$  of the PHS subunit, the corresponding values for L and S would be about 420,000 and 140,000, respectively.

The results of promoter probe cloning and nucleotide sequence analysis of the putative *phsA* promoter support the identity of the -10 and -35 regions and the transcriptional start point of *phsA*. Only a single start point was observed in the experiments illustrated in Fig. 5 and their replicates. It is possible that the *tsp* identified here is artifactual, but it is significant that the use of that start point identifies -10 and -35 regions with significant homology to the P2 promoter of the agarase gene (reference 5 and unpublished results). The -10 region, TCTCAT, of the *phsA* promoter showed more similarity to the -10 consensus sequence, TATAAT, of *E. coli*

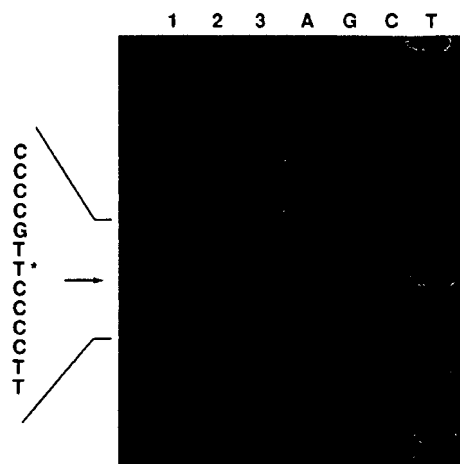


FIG. 5. Mapping of the 5' end of the *phsA* mRNA. A [ $\gamma$ - $^{32}$ P]ATP end-labeled 24-mer oligonucleotide (5'-GATCTCGGTCTCCCGCGTCACC-3') was annealed to *phsA* mRNA and extended with reverse transcriptase. The reaction products were separated on a sequencing gel with a sequencing ladder, generated by the use of the same primer, to determine the transcription start site of *phsA* mRNA. RNA templates were from *S. lividans* TK24 (lane 1), *S. antibioticus* (lane 2), and *S. lividans* transformed with pIJ2501 (lane 3). The arrow indicates the primer extension product that corresponds to initiation from *phsA*. The sequence on the left is the DNA region around the apparent transcription start site for *phsA*, indicated by an asterisk. Although the band corresponding to the extension product obtained with RNA from *S. antibioticus* is faint in the reproduction shown here (lane 2), it was clearly visible on the original autoradiograms.

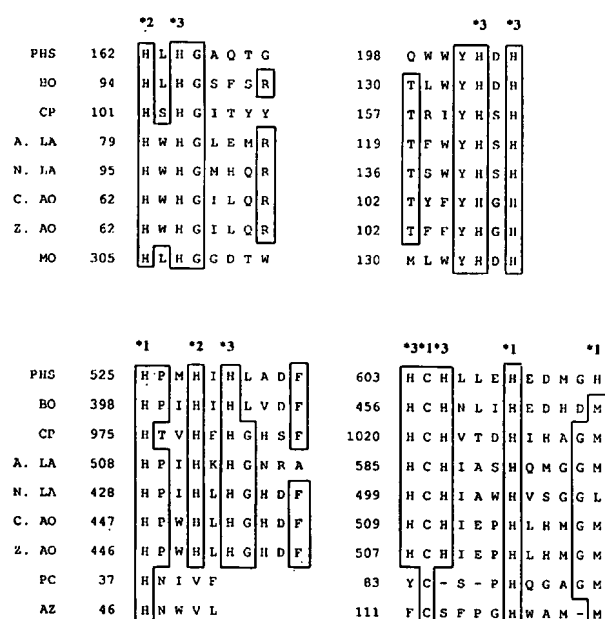


FIG. 6. Alignment of the putative copper-binding motifs of PHS, several blue copper proteins, and manganese-oxidizing protein. Sequence identities between PHS and the other blue copper proteins and manganese-oxidizing protein are boxed. The numbers to the left of the motifs denote the positions in the corresponding protein sequences. The amino acid residues corresponding to potential copper binding sites of three recognized types (40) are shown as follows: type I, \*1; type II, \*2; type III, \*3. Dashes represent gaps introduced to maximize the similarity. Protein sequences were from the following sources: BO, *Myrothecium verrucaria* bilirubin oxidase (32); CP, human ceruloplasmin (33); A. LA, *Aspergillus nidulans* laccase (1); N. LA, *Neurospora crassa* laccase (14); C. AO, cucumber ascorbate oxidase (44); Z. AO, zucchini ascorbate oxidase (40); PC, polar plastoquinone (43); AZ, *Alcaligenes denitrificans* azurin (43); MO, *Leptothrix discophora* manganese-oxidizing protein (8).

(15, 16) than to the  $-10$  consensus sequence, TAGGAT, of *Streptomyces* promoters (48). The  $-35$  region of the putative *phsA* promoter was not strikingly similar to the  $-35$  consensus sequence of either *E. coli* or *Streptomyces* promoters (Fig. 3) (see references 15, 16, and 48). Overall, the *phsA* promoter is not strongly homologous to any promoters for other antibiotic genes from *Streptomyces* spp. (48). However, recent studies do suggest similarities to the P2 promoter of the agarase gene (*dagA*) from *Streptomyces coelicolor* (reference 5 and unpublished data). Preliminary data also suggest that the *phsA* promoter is recognized by an alternative  $\sigma$  factor,  $\sigma^E$  (34). The role of this  $\sigma$  factor in *S. antibioticus* will be described in a subsequent publication.

One noteworthy feature of the *phsA* sequence is the presence of several sets of direct and inverted repeats near the promoter region (Fig. 3). This is especially interesting since some direct and inverted repeat sequences have been reported to be involved in the regulation of gene expression in streptomycetes. For example, repeated sequences have been implicated in the catabolite control of *Streptomyces* genes, including the chitinase genes of *Streptomyces plicatus* (9), the *galP1* promoter of the galactose operon of *S. lividans* (41), and  $\alpha$ -amylase promoters of *Streptomyces limosus* (51). None of the *phsA* direct or inverted repeat sequences is strikingly similar to the repeat sequences in the studies described above. However, it is possible that repeat motifs are a common feature of the regions involved in catabolite repression of streptomycete genes. The *phsA* sequence also contains two TNTNAN elements,

located within the  $-10$  region and upstream of the *phsA* promoter region (Fig. 3). TNTNAN hexamers were suggested to play a role in *galP1* regulation in *S. lividans* (41).

The predicted amino acid sequence of the PHS subunit resembles that of proteins belonging to the blue copper protein family. Like most members of this group (32), the sequence of the PHS subunit contains four consensus domains (1 to 4) that are presumed to bind the copper ligands (Fig. 6). Domains 1 and 2 are located at the N-terminal portion of the protein, whereas domains 3 and 4 are nearer the C terminus. Even though manganese-oxidizing protein does not belong to the blue copper protein family, similarities were observed in the copper binding domains 1 and 2 between PHS and the manganese-oxidizing protein (Fig. 6). In spite of the diverse distribution of these proteins and their utilization of very different substrates, they all use molecular oxygen in the reactions they catalyze. Although the active sites of these enzymes have not been characterized, the conservation of the copper binding sites strongly suggests their involvement in substrate recognition and catalysis.

In this study, we provided evidence for the regulation of the *phsA* promoter by catabolite repression. Thus, growth of *S. antibioticus* containing the cloned *phsA* promoter on glucose or glucose plus galactose led to a significant inhibition of *xylE* expression from pJSE935 as compared with that of cultures grown on galactose alone (Fig. 7). An identical pattern of inhibition was observed for PHS activity in the same cultures. There are several important implications of this result. First, the data strongly suggest that the sequences required for catabolite repression are contained within the *BsrBI* fragment cloned in pJSE935. This fragment lacks the repeats 1, 2, and 8 of Fig. 3. Thus, those sequences are presumably not required for catabolite repression. Second, it is clear that whatever the mechanism of catabolite repression in *S. antibioticus*, the relevant machinery can act simultaneously on the endogenous *phs* promoter and on the cloned sequence, since PHS activity parallels *xylE* activity in the experiments illustrated in Fig. 7. With regard to that mechanism, we presented evidence above that it may not involve a simple interaction between an operator and

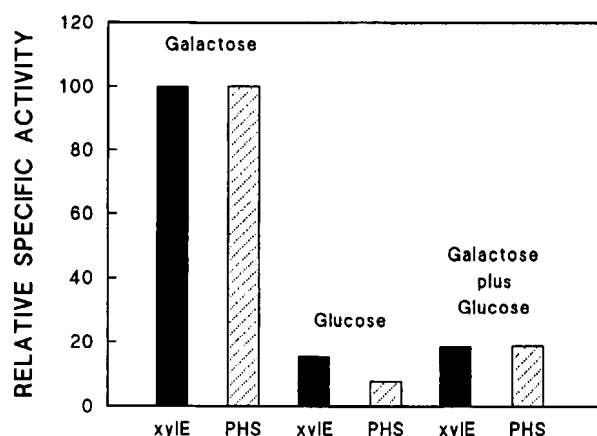


FIG. 7. Effects of carbon source on the expression of the *phsA* promoter in *S. antibioticus*. Transformants containing pJSE935 were grown on galactose, glucose, or glucose plus galactose as described in Materials and Methods. The figures show the results of catechol dioxygenase and PHS assays of extracts of mycelium harvested 12 h after inoculation. Results represent the averages of three replicates. The values obtained for extracts grown on galactose ( $42.4 \pm 2.1$  mU/mg of protein for catechol dioxygenase and  $75.6 \pm 5.3$  U/mg of protein for PHS) were arbitrarily set at 100 for purposes of presentation.

a repressor as it was not possible to release expression of the endogenous *phs* gene from catabolite repression by cloning the disrupted gene at high copy. It is possible, of course, that the repressor binding site involves sequences that were disrupted by the insertion of the *Xba*I linker. It should be possible to distinguish between these possibilities and to learn more about the mechanism of catabolite repression of *phsA* expression by gel mobility shift assays.

#### ACKNOWLEDGMENTS

We thank Mervyn Bibb for the FRAME analysis. This work was supported by a grant from Emory University.

#### REFERENCES

- Aramayo, R., and W. E. Timberlake. 1990. Sequence and molecular structure of the *Aspergillus nidulans* *ya* (laccase I) gene. *Nucleic Acids Res.* 18:3415.
- Barry, C. E., P. G. Nayar, and T. P. Begley. 1989. Phenoxazinone synthase: mechanism for the formation of the phenoxazinone chromophore of actinomycin. *Biochemistry* 28:6323-6333.
- Bibb, M. J., P. R. Findlay, and M. W. Johnson. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* 30:157-166.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1523.
- Buttner, M. J., A. M. Smith, and M. J. Bibb. 1988. At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). *Cell* 52:599-607.
- Choy, H. A., and G. H. Jones. 1981. Phenoxazinone synthase from *Streptomyces antibioticus*: purification of the large and small enzyme forms. *Arch. Biochem. Biophys.* 211:55-65.
- Clayton, T. M., and M. J. Bibb. 1990. *Streptomyces* promoter-probe plasmids that utilize the *xylE* gene of *Pseudomonas putida*. *Nucleic Acids Res.* 18:1077.
- Corstjens, P. L. A. M. 1993. Ph.D. thesis. Rijksuniversiteit, Leiden, The Netherlands.
- Delic, I., P. Robbins, and J. Westpheling. 1992. Direct repeat sequences are implicated in the regulation of two *Streptomyces* chitinase promoters that are subject to carbon catabolite control. *Proc. Natl. Acad. Sci. USA* 89:1885-1889.
- Favre, D. 1992. Improved phenol-based method for the isolation of DNA fragments from low melting temperature agarose gels. *BioTechniques* 13:24-26.
- Fawaz, F., and G. H. Jones. 1988. Actinomycin synthesis in *Streptomyces antibioticus*: purification and properties of 3-hydroxyanthranilate 4-methyltransferase. *J. Biol. Chem.* 263:4602-4606.
- Freeman, J. C., P. G. Nayar, T. P. Begley, and J. J. Villafranca. 1993. Stoichiometry and spectroscopic identity of copper centers in phenoxazinone synthase: a new addition to the blue copper oxidase family. *Biochemistry* 32:4826-4830.
- Gallo, M., and E. Katz. 1972. Regulation of secondary metabolite biosynthesis: catabolite repression of phenoxazinone synthase and actinomycin formation by glucose. *J. Bacteriol.* 109:659-667.
- Germann, U. A., G. Muller, P. E. Hunziker, and K. Lerch. 1988. Characterization of two allelic forms of *Neurospora crassa* laccase. *J. Biol. Chem.* 263:885-896.
- Harley, C. B., and R. P. Reynolds. 1987. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* 15:2343-2361.
- Hawley, D. K., and W. R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* 11:2237-2255.
- Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schrempf. 1985. Genetic manipulation of streptomycetes: a laboratory manual. The John Innes Foundation, Norwich, United Kingdom.
- Hopwood, D. A., T. Kieser, H. M. Wright, and M. J. Bibb. 1983. Plasmids, recombination and chromosome mapping in *Streptomyces lividans*. *J. Gen. Microbiol.* 129:2257-2269.
- Ingram, C., M. Brawner, P. Youngman, and J. Westpheling. 1989. *xylE* functions as an efficient reporter gene in *Streptomyces* spp.: use for the study of *galP1*, a catabolite-controlled promoter. *J. Bacteriol.* 171:6617-6624.
- Jones, G. H. 1985. Regulation of phenoxazinone synthase expression in *Streptomyces antibioticus*. *J. Bacteriol.* 163:1215-1221.
- Jones, G. H. 1985. Regulation of actinomycin synthesis in *Streptomyces antibioticus*. *J. Nat. Prod. (Lloydia)* 49:981-987.
- Jones, G. H. 1987. Actinomycin synthesis in *Streptomyces antibioticus*: enzymatic conversion of 3-hydroxyanthranilic acid to 4-methyl-3-hydroxyanthranilic acid. *J. Bacteriol.* 169:5575-5578.
- Jones, G. H. 1993. Combined purification of actinomycin synthetase I and 3-hydroxyanthranilic acid 4-methyltransferase from *Streptomyces antibioticus*. *J. Biol. Chem.* 268:6831-6834.
- Jones, G. H., and D. A. Hopwood. 1984. Activation of phenoxazinone synthase expression in *Streptomyces lividans* by cloned DNA sequences from *Streptomyces antibioticus*. *J. Biol. Chem.* 259:14158-14164.
- Jones, G. H., and D. A. Hopwood. 1984. Molecular cloning and expression of the phenoxazinone synthase gene from *Streptomyces antibioticus*. *J. Biol. Chem.* 259:14151-14157.
- Katayama, C. 1990. Single-stranded DNA sequencing: a simplified protocol for single-stranded DNA rescue. *Strategies* 4:56-58.
- Katz, E., C. J. Thompson, and D. A. Hopwood. 1983. Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. *J. Gen. Microbiol.* 129:2703-2714.
- Katz, E., and H. Weissbach. 1962. Biosynthesis of the actinomycin chromophore; enzymatic conversion of 4-methyl-3-hydroxyanthranilic acid to actinocin. *J. Biol. Chem.* 237:882-886.
- Keller, U. 1987. Actinomycin synthetases; multifunctional enzymes responsible for the synthesis of the peptide chains of actinomycin. *J. Biol. Chem.* 262:5852-5856.
- Keller, U., H. Kleinkauf, and R. Zocher. 1984. 4-methyl 3-hydroxyanthranilic acid activating enzyme from actinomycin producing *Streptomyces chrysomallus*. *Biochemistry* 23:1479-1484.
- Keller, U., and W. Schlumbohm. 1992. Purification and characterization of actinomycin synthetase I, a 4-methyl-3-hydroxyanthranilic acid-AMP ligase from *Streptomyces chrysomallus*. *J. Biol. Chem.* 267:11745-11752.
- Koikeda, S., K. Ando, H. Kaji, T. Inoue, S. Murao, K. Takeuchi, and T. Samejima. 1993. Molecular cloning of the gene for bilirubin oxidase from *Mycobacterium verrucaria* and its expression in yeast. *J. Biol. Chem.* 268:18801-18809.
- Koschinsky, M. L., W. D. Funk, B. A. v. Oost, and R. T. A. MacGillivray. 1986. Complete cDNA sequence of human preceruloplasmin. *Proc. Natl. Acad. Sci. USA* 83:5086-5090.
- Lonetto, M., K. L. Brown, K. Rudd, and M. J. Buttner. 1994. Analysis of the *Streptomyces coelicolor* *sigE* gene reveals a new sub-family of bacterial RNA polymerase  $\sigma$  factors involved in the regulation of extracytoplasmic functions. *Proc. Natl. Acad. Sci. USA* 91:7573-7577.
- Luria, S. E., J. N. Adams, and R. C. Ting. 1960. Transduction of lactose utilizing ability among strains of *E. coli* and *S. dysenteriae* and the properties of the transducing phage particles. *Virology* 12:348-356.
- Lydiate, D. J., F. Malpartida, and D. A. Hopwood. 1985. The *Streptomyces* plasmid SCP2<sup>+</sup>: its functional analysis and development into useful cloning vectors. *Gene* 35:223-235.
- Madu, A. C., and G. H. Jones. 1989. Molecular cloning and *in vitro* expression of a silent phenoxazinone synthase gene from *Streptomyces lividans*. *Gene* 84:287-294.
- Malpartida, F., and D. A. Hopwood. 1984. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature (London)* 309:462-464.
- Marshall, R., B. G. Redfield, E. Katz, and H. Weissbach. 1968. Changes in phenoxazinone synthase activity during the growth cycle of *Streptomyces antibioticus*. *Arch. Biochem. Biophys.* 123:317-323.
- Masserschmidt, A., A. Rossi, R. Ladenstein, R. Huber, M. Bolognesi, G. Gatti, A. Marchesini, R. Petruzzelli, and A. Finazzi-Agro. 1989. X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini: analysis of the polypeptide fold and a model of the copper sites and ligands. *J. Mol. Biol.* 206:513-529.
- Mattern, S. G., M. E. Brawner, and J. Westpheling. 1993. Identification of a complex operator for *galP1*, the glucose-sensitive, galactose-dependent promoter of the *Streptomyces* galactose operon. *J. Bacteriol.* 175:1213-1220.
- Moran, C. P., Jr. 1990. Measuring gene expression in *Bacillus*, p. 267-293. In C. R. Harwood and S. M. Cutting (ed.), *Molecular biology methods for Bacillus*. John Wiley & Sons, Inc., New York.
- Norris, G. E., B. F. Anderson, and E. N. Baker. 1983. Structure of azurin from *Alcaligenes denitrificans* at 2.5 Å resolution. *J. Mol. Biol.* 165:501-521.
- Ohkawa, J., N. Okada, A. Shinmyo, and M. Takano. 1989. Primary structure of cucumber (*Cucumis sativus*) ascorbate oxidase deduced from cDNA sequence: homology with blue copper proteins and tissue-specific expression. *Proc. Natl. Acad. Sci. USA* 86:1239-1243.
- Sala-Trepat, J. M., and W. C. Evans. 1971. The meta cleavage of catechol by *Azotobacter* species 4-oxalocrotonate pathway. *Eur. J. Biochem.* 20:400-413.
- Sambrook, J., E. F. Fritsch, and T. E. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Strohl, W. R. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoter. *Nucleic Acids Res.* 20:961-974.
- Thompson, C. J., J. M. Ward, and D. A. Hopwood. 1982. Cloning of antibiotic resistance and nutritional genes in streptomycetes. *J. Bacteriol.* 151:668-677.
- Troust, T., and E. Katz. 1979. Phenoxazinone biosynthesis: accumulation of

- a precursor, 4-methyl-3-hydroxyanthranilic acid, by mutants of *Streptomyces parvulus*. J. Bacteriol. 169:5575-5578.
51. Virolle, M.-J., and J. Gagnat. 1994. Sequences involved in growth-phase-dependent expression and glucose repression of a *Streptomyces*  $\alpha$ -amylase gene. Microbiology 140:1059-1067.
52. Waksman, S. A. 1968. Actinomycin: nature, formation and activities. John Wiley & Sons, Inc., New York.
53. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.
54. Zukowski, M. M., D. F. Gaffney, D. Speck, M. Kauffmann, A. Findeli, A. Wisecup, and J.-P. Lecocq. 1983. Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. Proc. Natl. Acad. Sci. USA 80:1101-1105.

## The Sequence of Spacers between the Consensus Sequences Modulates the Strength of Prokaryotic Promoters

PETER RUHDAL JENSEN\* AND KARIN HAMMER

Department of Microbiology, Technical University of Denmark, DK-2800 Lyngby, Denmark

Received 7 July 1997/Accepted 21 October 1997

We constructed a library of synthetic promoters for *Lactococcus lactis* in which the known consensus sequences were kept constant while the sequences of the separating spacers were randomized. The library consists of 38 promoters which differ in strength from 0.3 up to more than 2,000 relative units, the latter among the strongest promoters known for this organism. The ranking of the promoter activities was somewhat different when assayed in *Escherichia coli*, but the promoters are efficient for modulating gene expression in this bacterium as well. DNA sequencing revealed that the weaker promoters (which had activities below 5 relative units) all had changes either in the consensus sequences or in the length of the spacer between the -35 and -10 sequences. The promoters in which those features were conserved had activities from 5 to 2,050 U, which shows that by randomizing the spacers, at least a 400-fold change in activity can be obtained. Interestingly, the entire range of promoter activities is covered in small steps of activity increase, which makes these promoters very suitable for quantitative physiological studies and for fine-tuning of gene expression in industrial bioreactors and cell factories.

Metabolic engineering has promising perspectives with respect to improving the properties and performances of microorganisms used as industrial bioreactors, as cell factories, and in food fermentations. The importance of tuning gene expression in this context, i.e., to perform metabolic optimization rather than massive overexpression or gene inactivation, is now far more appreciated. However, the more subtle approach of metabolic optimization is hampered by the lack of proper expression systems for tuning gene expression in many microorganisms. Also, the fundamental understanding of a biological system through metabolic control analysis (5, 10) requires the tuning of enzyme activities in order to calculate the so-called control coefficients. For some organisms, expression systems that allow for changing gene expression for scientific purposes and for a limited set of experimental conditions have been developed. Thus, for *Escherichia coli*, the *lac* system, the *cl*-regulated lambda  $p_R/p_L$ , and many derivatives of these systems have been widely applied, and such systems have also been adapted for use in other organisms (for a recent review, see reference 12). With respect to changing steady-state gene expression, these systems can sometimes be difficult to apply, particularly when it comes to changing gene expression on an industrial scale. Besides, in most food fermentation processes, the addition of chemicals as inducers of gene expression or the changing of other process parameters is not acceptable; in such cases, there are virtually no expression systems available for tuning gene expression and thus for performing accurate metabolic optimization.

Lactic acid bacteria are widely used in food fermentation, e.g., cheese and yoghurt production, but besides lactic acid, these bacteria excrete a spectrum of organic compounds. Some of these are desirable with respect to the development of texture and flavors or for bioconservation purposes, and some are undesirable for similar or different reasons. The lactic acid bacteria are therefore obvious candidates for attempts to op-

timize the pattern of formation of these compounds for specific applications. But the experimental tools for manipulating gene expression are not well developed for these bacteria. An exception is the nisin-inducible system, developed recently by de Ruyter et al. (2). This system appears to be well suited for inducing gene expression in *Lactococcus lactis* by adding the antibiotic nisin (which is accepted as a food additive). A question that perhaps needs to be addressed in this context is whether the nisin expression system is also suitable for achieving a steady level of gene expression. In addition, for effective metabolic optimization, it is often necessary to optimize the expression of a number of genes, which is not feasible with the systems developed so far.

Here we describe a method for tuning steady-state gene expression in *L. lactis*. We overcome many of the limitations discussed above by using libraries of synthetic promoters which cover a wide range of promoter activities and show that the strength of prokaryotic promoters can be modulated by randomizing the spacer sequences that separates the consensus sequences. The system is food grade and well suited for use in industrial bioreactors and food fermentation processes. In addition, the system should be applicable to a broad range of biological systems. (Potential commercial users should be aware that the approach for obtaining the synthetic promoters, as well as the promoter sequences, were filed for patent worldwide [7a]).

### MATERIALS AND METHODS

**Bacterial strains and plasmids.** The *E. coli* K-12 strain BOE270 (1) is highly competent with respect to transformation and was derived from strain MT102, which in turn is an *hsdR* derivative of strain MC1000 [*araD139*  $\Delta$ (*ara-leu*)7679 *galU galK*  $\Delta$ (*lac*)174 *rpsL thi-1* (1a)]. BOE270 was used for studying promoter activities in *E. coli* as well as for cloning purposes and propagation of plasmid DNA in *E. coli*. The plasmid-free *L. lactis* subsp. *cremonis* strain MG1363, which does not express  $\beta$ -galactosidase activity (4), was used for studying promoter activities in *L. lactis*.

The promoter cloning vector pAK80 (7) was used for cloning the synthetic promoters DNA fragments. pAK80 is a shuttle vector for *L. lactis* and *E. coli*, conferring erythromycin resistance to the host cells. The vector carries the promoterless *lacL* and *lacM* genes from *Leuconostoc lactis* (which codes for  $\beta$ -galactosidase enzyme activity). It contains a multiple cloning site for the insertion of DNA fragments harboring putative promoter signals, just upstream

\* Corresponding author. Mailing address: Department of Microbiology, Technical University of Denmark, Building 301, DK-2800 Lyngby, Denmark. Phone: 45 45252510. Fax: 45 45932809. E-mail: prj@im.dtu.dk.

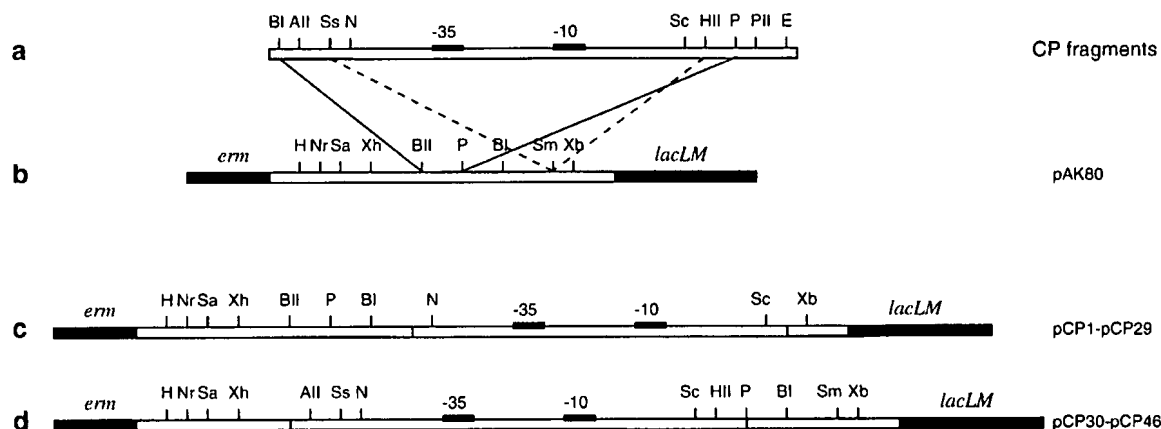


FIG. 1. Strategies used for cloning synthetic promoter fragments into the promoter cloning vector pAK80. (a) Double-stranded DNA fragments carrying putative promoter activities. (b) Restriction map and schematic representation of the relevant parts of the promoter cloning vector. The stippled and solid lines show the strategies used for cloning pCP1 through pCP29 and pCP30 through pCP46, respectively. (c) Restriction map of clones pCP1 through pCP29. (d) Restriction map of clones pCP30 through pCP46. Note that a number of clones have been subject to cloning artifacts and thus may have a slightly different restriction map. BI, *Bam*HI; All, *Afl*III; Ss, *Ssp*I; N, *Nsi*I (*Pst*I compatible); Nr, *Nru*I; Sc, *Sc*I; HII, *Hinc*II; P, *Pst*I; PII, *Pvu*II; E, *Eco*RI; Sa, *Sac*I; Xh, *Xho*I; BII, *Bgl*II; Sm, *Sma*I; Xb, *Xba*I (not drawn to scale).

the promoterless *lacL* and *lacM* genes from *Leuconostoc lactis*. Together, the *lacL* and *lacM* genes codes for a  $\beta$ -galactosidase.

**Enzymes.** Restriction enzymes, Klenow DNA polymerase, calf intestine phosphatase, and T4 DNA ligase were obtained from and used as recommended by Pharmacia and New England Biolabs.

**Oligonucleotides.** Oligonucleotides were obtained from Hobolth DNA Synthesis (Hillerød, Denmark).

**Second-DNA-strand synthesis.** The single-stranded promoter oligonucleotides were converted to double-stranded DNA, using a 10-bp oligonucleotide (5'-CC GAATTCAG) complementary to the 3' end of the promoter oligonucleotide as primer for the second-strand synthesis by the Klenow fragment of DNA polymerase I.

**Cloning of synthetic DNA fragments into the promoter cloning vector pAK80.** Two different cloning strategies were used (Fig. 1). In strategy A, the mixture of DNA fragments was digested with two restriction enzymes, *Hinc*II and *Ssp*I, and pAK80 was digested with *Sma*I. In strategy B, the mixture of DNA fragments was digested with two restriction enzymes, *Bam*HI and *Pst*I, and pAK80 was digested with *Bgl*II and *Pst*I. In both strategies, the promoter fragments were then ligated to the compatible vector fragments. The ligation mixtures were then transformed into  $Ca^{2+}$ -competent cells (13) by using a standard transformation procedure (13), and the transformation mixture were plated (at 30°C) on LB plates containing erythromycin (200  $\mu$ g/ml) and 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal; 100  $\mu$ g/ml). A total of 150 erythromycin-resistant transformants were obtained; all were white initially, but after prolonged incubation (up to 2 weeks at 4°C), a number had become blue to various extents. Later, we discovered that the development of blue color from *E. coli* colonies (but not *L. lactis* colonies) expressing *lacLM* is greatly enhanced by adding 1% glycerol to the transformation plates (data not shown). Plasmids were isolated from these blue colonies, and it was confirmed by restriction enzyme analysis that most of these clones had promoter fragments inserted in the multiple cloning site of pAK80, in the orientation that would direct transcription into the  $\beta$ -galactosidase gene (*lacLM*). The 46 colonies isolated had become blue to various extents; 29 from cloning strategy A (containing plasmids pCP1 through pCP29) and 17 from strategy B (containing plasmids pCP30 through pCP46) were picked for further analysis. The two weakest promoter clones, pCP31 and pCP43, did not contain a promoter fragment, and four promoter clones, pCP18, pCP19, pCP33, and pCP44, turned out to be identical to pCP27, pCP22, pCP35, and pCP45, respectively. Indeed, the activities of these sets were almost identical, which also demonstrates the reproducibility of the assay used here. The chances that two identical sequences would have arisen by coincidence during the oligonucleotide synthesis is of course negligible, and these four clones must therefore be the result of a cell division that took place after the plasmids were transformed but before the cells were plated.

**Transformation of *L. lactis*.** Cells of *L. lactis* subsp. *cremoris* MG1363 (4) were made competent by growth overnight in GM17 medium containing 2% glycine as described by Holo and Ness (6). Plasmid DNA from the 46 clones described above was then transformed into these cells by electroporation (6). The cells were allowed to regenerate in SGM17 medium for 2 h and then plated on SR plates containing erythromycin (2  $\mu$ g/ml) and X-Gal (100  $\mu$ g/ml).

**$\beta$ -Galactosidase assay.** The assay was done as described by Miller (14) and modified by Israelsen et al. (7). Cultures carrying the plasmid derivatives of pAK80 were grown in rich medium overnight at 30°C. The medium used for

*L. lactis* was M17 medium supplemented with erythromycin (2  $\mu$ g/ml) and 1% glucose; for *E. coli*, LB medium supplemented with erythromycin (200  $\mu$ g/ml) was used. The results presented are averages of measurements of the activities of at least three individual cultures of each clone. The standard errors were less than 30% for *E. coli* activities and less than 20% for *L. lactis* activities. Aliquots of 25 to 100  $\mu$ l of the cultures were used in the  $\beta$ -galactosidase assay except in the case of the weakest promoter clones, where up to 2 ml of culture was concentrated and used in the assay.

## RESULTS

The purpose of this work was to generate a library of synthetic constitutive promoters as a tool for genetic engineering of *L. lactis*. The promoters should cover a wide range of promoter activities, in small steps of activity changes, so that they would be applicable to quantitative physiological studies and for metabolic optimization. The following strategy was used: (i) design and synthesize a degenerated oligonucleotide sequence that encodes consensus sequences for *L. lactis* promoters, separated by spacers of random sequences; (ii) convert this mixture of oligonucleotides to double-stranded DNA fragments, using DNA polymerase and a short oligonucleotide primer complementary to the 3' end of the degenerated oligonucleotide; and (iii) clone this mixture of DNA fragments into a promoter probing vector. The idea behind this strategy is that even though the consensus sequences should be important elements of an efficient promoter, the context in which the consensus sequences are located may modulate the strength of the promoters to some extent.

**Design and construction of synthetic promoters for *L. lactis*.** A considerable number of promoters have been cloned and sequenced from *L. lactis* (see the review by de Vos and Simons [3]). From these data, we extracted extended consensus sequence motifs for *L. lactis* promoters (Fig. 2A). The Pribnow box or the -10 sequence TATAAT and the -35 sequence TTGACA, known to be present in many prokaryotic promoters, are also well conserved for *L. lactis*. In addition, the sequence TG is often found 1 bp upstream of the -10 sequence; it is also possible to determine a consensus sequence for the 4 bp immediately upstream of the -35 motif, ATTC. Nilsson and Johansen (16) found well-conserved sequences among promoters of the rRNA operons: AGTTT at position -44 and GTACTGTT at positions +1 to +8. In addition to these mo-



Promoter		Spec. $\beta$ -gal activity	
		<i>L. lactis</i> <i>E. coli</i>	
<b>Oligoseq</b>	<b>CATNNNNNAGTTTATTCTTGACANNNNNNNNNNNNTGRTATAATANNWNAGTACTGTT</b>		
CP24	CATGGGTAAAGTTTATTCTTGCACACTATCTGGGCCCCGATGGTATAATAAGTACTACTGTT	0.3	3.1
CP18=CP27	CATTTTGACAGTTTATTCTTGACATTGTGTGCTTCGGGTG-TATAATA-CTAAGTACTGTT	0.7	0.2
CP37	CATCATTAAGTTTATTCTTGCACATTGGCCGGAATTGTTG-TATAATACCTTAGTACTGTT	2.2	16
CP17	CATTCTGGAAGTTTATTCTTGAC-CGCTCAGTATGCAGTGGTATAATAGTACAGTACTGTT	2.7	10
CP2	CATTTGCTAGTTTATTCTTGACATGAAGCGTGCCATAGGTATAATTACTTGGTACTGTT	4.9	2.8
CP4	GATGTTTAAAGTTTATTCTTGACACCGTATCGTGGCGGTGATATAATCGGGATCCTTAAGA	5.1	1.3
CP44=CP45	CATCGGTAAGTTTATTCTTGACA-ATTAAAGTAGAGCTGATATAATAGTTCAGTACTGTT	12	34
CP1	CATACCGGAGTTTATTCTTGACAGTTCCACCTCGGGTGTATATAATATCTCAGTACTGTT	34	3.1
CP19=CP22	CATCGCTAGTTT-TTCTTGACAGGAGGATCCGGGTGTATATAATA-GTTAGTACTGTT	58	3.3
CP34	CATCGCGAAGTTTATTCTTGACACACCGCAGAACTTGGTATAATACACAGTACTGTT	59	8.4
CP20	CATGGGTAGTTTATTCTTGACAGTGGCGCGGGGGCTGATATCATAGCAGTACTATT	60	22
CP11	CATAAGTAGTTTATTCTTGACCGGACGCCCCCTTGTATATAATAAGT-AGTACTGTT	69	1.4
CP26	CATTCTACAGTTTATTCTTGACATTGCACGTGCCCCCTGGTATAATAACTATACATGCAT	72	10
CP3	CATCCTGTAGTTTATTCTTGACACAAAGTCGTTAGCTGGTATAATAGGAGAGTACTGTT	74	0.9
CP14	CATGACGAGTTTATTCTTGACACAGGTATGGACTTATGATATAATAAACAGTACTGTT	81	0.3
CP13	CATGCTTTACTTTATTCTTGACAAAACCCAGCTTTGGTATAATACGTGAGAACTGTT	84	1.0
CP40	CATAGAACAGTTTATTCTTGACATTGAATAAGAGGCTGATATAATAGC-CAGTACTGTT	104	19
CP8	CATTCTTAGTTTATTCTTGACAAACGATTGAGGACTGATATAATAGGTAGTACTGTT	146	1.2
CP28	CATGGGGCGTTTATTCTTGACACCGGCGAGCAGCCTGGTATAATAATAGTACTGTT	161	0.9
CP10	CATGGCTTAGTTTATTCTTGACAGGGTAGTATCACTGGTATAATAGGACAGTACTGTT	181	2.5
CP32	CATACGGGAGTTTATTCTTGACATATTGCGGGTGTGGTATAATAACTTAGTACTGTT	214	60
CP30	CATAGCAGAGTTTATTCTTGACAGTATTGGGTTACTTGGTATAATAGTTAGTACTGTT	228	42
CP9a	CATAGTCTAGTTTATTCTTGACACCGGCTCCATTGGCTGGTATAATAATTAGTACTGTT	243	56
CP38	CATAGAGAAGTTTATTCTTGACAGCTAACTTGGCCCTTGTATATAATACATGAGTACTGTT	244	92
CP46	CATGATGTAGTTTATTCTTGACACTGAGAGGGCCCTTGTATATAATAGTTAGTACTGTT	256	33
CP23	CATGTAGGAGTTTATTCTTGACAGATTAGTTAGGGGGTGGTATAATACTCAGTACTGTT	257	2.6
CP39	CATTTCGAAGTTTATTCTTGACAGTACGTTTACCATTGATATAATAGTATAGTACTGTT	266	0.5
CP33=CP35	CATGTTGGAGTTTATTCTTGACATACAACTACTGCGATGATATAATAGGTAGTACTGTT	298	7.0
CP15	CATTACGTAGTTTATTCTTGACAGAACTACGATTCGCTGGTATAATATACAGTACTGTT	322	1.5
CP29	CATCGCTAAG-TTATTCTTGACATCTCAGGGGGGACGTGGTATAATAACTGAGTACTGTT	399	1.4
CP12b	CATATACAAGTTTATTCTTGACACTAGTCGGCCAAATGATATAATACCTGAGTACTGTT	419	101
CP41	CATCCGCAAGTTTATTCTTGACAGCTGAATGTAGACGTGGTATAATAGTTAAGTACTGTT	624	18
CP16	CATTGTGTAGTTTATTCTTGACAGCTATGAGTCAATTTGGTATAATA--ACAGTACTCAG	627	0.3
CP42	CATTTCGAAGTTTATTCTTGACACCTGAGATGAGGCGTGATATAATAAATAGTACTGTT	680	7.2
CP7	TATGGCGTAGTTTATTCTTGACATGACGAGACAGGTGGTATAATGGGTCTAGATTAGG	1021	134
CP6	CATTCGAGAGTTTATTCTTGACAGATATTTCCGGATGATATAATAACTGAGTACTGTT	1876	280
CP25	C-TTTCGAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTGTT	2050	528

FIG. 4. Sequence of the area from positions -52 to +8 (relative to the putative transcription initiation site) of the synthetic promoter clones pCP1 through pCP46. The clones are ordered according to strength. Matches to the oligonucleotide consensus sequence (given at the top) are in boldface. Errors in the -35 or -10 consensus sequence and deletions in the spacer between these sequences are underlined. Two clones, CP9 and CP12, had two promoter fragments inserted in tandem, a (upstream fragment) and b (downstream fragment). In these cases, only one of the two tandem promoters was perfect; data for these promoters are shown.  $\beta$ -gal,  $\beta$ -galactosidase.

**Regulation of promoter activities.** The synthetic CP promoters were designed to be constitutive. To test this experimentally, the expression in exponential growth phase and stationary growth phase was measured for a selection of the promoter clones. We found that the specific activity of  $\beta$ -galactosidase was two- to fourfold higher in the stationary-phase cultures than in the exponential-phase cultures (data not shown). However, the copy number of the vector used in these studies has been shown to increase approximately threefold in the stationary phase (11), which demonstrates that the CP promoters are indeed quite close to being constitutive under these conditions.

**Activities of the synthetic promoters in *E. coli*.** Another interesting point is whether the promoters are functional in other organisms, and if so, whether the relative strength of the promoters would be dependent on the organism. As described above, the promoter cloning vector, pAK80, that we used here for construction of the synthetic promoters also replicates in *E. coli*; indeed, the promoter clones were first isolated in *E. coli*. We could therefore measure the activities of the synthetic promoters also in *E. coli* (Fig. 5). The promoter strength was also highly variable for the individual promoters in this organism, and we found that the promoters covered activities from 0.2 to 500 Miller units. In this case also, the activity increased in small steps.

The absolute values of  $\beta$ -galactosidase units measured in *E. coli* were lower on average compared to *L. lactis*; this was probably a consequence of a low efficiency of translation of the *lacL* and *lacM* genes in *E. coli*, since these genes and their ribosome binding sites originate from the gram-positive bacterium *Leuconostoc mesenteroides*. When some of the strongest promoters were cloned into a promoter cloning vector designed for *E. coli*, the promoters turned out to be quite strong (data not shown).

Figure 6 shows a plot of activity of the CP promoters in

*L. lactis* and *E. coli*. The strengths of the individual CP promoters in the two organisms correlate somewhat but not very well: some promoters which were quite strong in *L. lactis* were relatively weak in *E. coli*, and vice versa. Moreover, the pattern that we observed in *L. lactis*, i.e., that the relatively strong promoters were the perfect ones, did not hold true for *E. coli*: here the promoters which had either an error in the consensus sequence or a shorter spacer were relatively strong.

## DISCUSSION

We have constructed a library of synthetic promoters that differ in strength over 3 to 4 logs of activity, and this range of activity is covered by small steps of activity increase. Moreover, some of the promoters that resulted from this random approach turned out to be quite strong.

The fact that the library of promoters covered such a wide range of activities was somewhat surprising to us; the underlying idea behind the construction of the CP promoters was that the context of the consensus sequences (the spacers) would play a role in modulating the strength of a promoter, rather than changing the activity over several logs of activity. Indeed, much of that variation (below 5 Miller units) was probably a consequence of the accidental introduction of mutations in the consensus sequences and in the length of the spacer regions. In contrast, the strong promoters in *L. lactis* (those having activities higher than 100 Miller units) were all perfect with respect to the consensus sequence and spacer length. But even when we confine our analysis to these promoter clones, we find 400-fold variation in promoter activity, still in small steps of activity increase, which demonstrates that the context in which the consensus sequences are embedded (i.e., the spacers) clearly is important for promoter strength.

The ranking of the promoters depended on the organism in



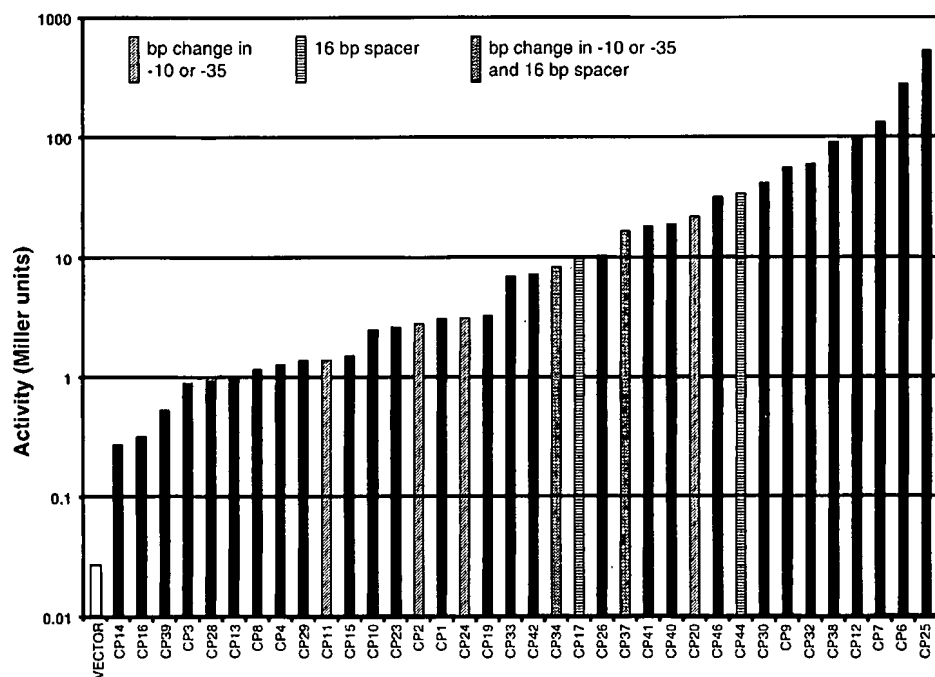


FIG. 5.  $\beta$ -Galactosidase activities of the CP promoters in *E. coli*. The promoter activities were assayed from the expression of a reporter gene (*lacLM*) encoding  $\beta$ -galactosidase transcribed from the different synthetic promoter clones on the promoter cloning vector pAK80. The patterns of the data points indicate which promoter clones contained errors in either the  $-35$  or the  $-10$  consensus sequence or in the length of the spacer between these sequences. See text for further details.

which they were measured, possibly because the  $\sigma$  factor-RNA polymerase complexes that recognize these promoters have different structures in the two organisms due to differences in amino acid sequences. The fact that *E. coli* accepted some of the less perfect CP promoters as relatively strong promoters could indicate that *E. coli* is more promiscuous with respect to promoter structure than *L. lactis*. This makes some sense considering the composition of the *L. lactis* genome: the AT content is 65%, which is much closer to the base composition of the  $-35$  and  $-10$  consensus sequences. These sequences are therefore more likely to accidentally occur in *L. lactis*, and a stricter requirement for promoter sequences might therefore be expected for this organism.

The process of transcription initiation consists of several events (reviewed in reference 17). First, recognition and binding of the  $\sigma$  factor-RNA polymerase complex to the promoter region takes place (closed complex formation). Subsequently, there is local melting of the DNA double helix (open complex formation), possibly assisted by local negative DNA supercoiling. Finally, the binding between the  $\sigma$  factor-RNA polymerase complex and the promoter area must dissociate and clear the promoter area, so that another initiation complex may form. From this model, it is clear that efficient binding between the  $\sigma$  factor-RNA polymerase complex and the promoter area does not guarantee a strong promoter; promoter strength must be a compromise between binding, melting, and clearance, and probably other factors as well.

What then controls the strength of the individual synthetic promoters presented here? It does not appear that any additional conserved sequence motifs have been generated among the strongest promoters. Rather, it seems that the overall three-dimensional structure which arises from a particular nucleotide sequence could be important.

The method presented here for tuning gene expression in

the living cell has both advantages and disadvantages compared to the methods that would use an inducible expression system such as the *lac* promoter. A disadvantage is that instead of only one genetic construct, perhaps three to four constructs have to be made. On the other hand, the constructs are made

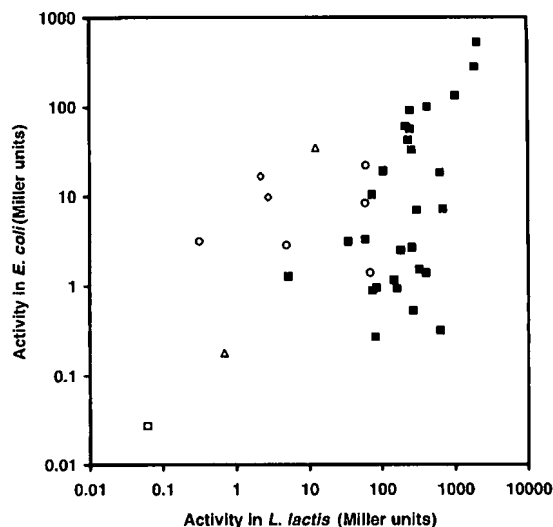


FIG. 6. Correlation between promoter activities in *L. lactis* and *E. coli*. The promoter activities measured in *E. coli* (from Fig. 5) were plotted as a function of the promoter activities measured in *L. lactis* (from Fig. 3). The symbols indicate errors in either the  $-35$  or  $-10$  sequence (solid circles), a 16-bp spacer (triangles), or promoters with both of these errors (diamonds). The open square represents the vector clone.

in parallel, so that the amount of work should not be proportional to the number of constructs. The inducible systems have the advantage that gene expression can be turned on at the proper time during a fermentation, which is sometimes essential (for instance, when the product is toxic to the host cell). The work presented here was aimed at generating a library of constitutive promoters, for achieving a constant level of gene expression throughout the growth of a culture. We are currently working on synthetic inducible promoters in which a regulatory motif has been added. This should allow us to generate libraries of promoters, which differ in basal expression level and can be induced to various extents, by changing a fermentation parameter (i.e., temperature, pH, or salt concentration) or by adding a specific inducer.

The system presented here also has advantages. One is that it is easier to attain a steady expression level of the enzyme in question, which is often quite difficult with inducible systems such as the *lac* system (8). With the method presented here, once the optimal expression level of the enzyme has been determined, the engineered strain is ready to use directly in the fermentation process.

An important feature of the system described here, in a longer perspective, is the possibility to simultaneously modulate, to different extents, the expression of several individual genes or operons located at various positions of the genome in the same strain. Metabolic control analysis (5, 10) showed that in theory, flux and concentration control can be shared among several enzymes in a pathway, and experimental determinations of flux control have often showed that control seems to be distributed over many enzymes in the living cell (9, 15, 18, 19, 22, 23): in most cases, there may not be such a thing as a rate-limiting step, and even if one finds a step that has a measurable control, the control will often disappear relatively quickly as the enzyme is being overexpressed. Since the sum of flux control must equal unity, this then means that flux control has been shifted to other steps in the pathway. In summary, in order to increase a given flux in a living cell, it may thus be necessary to (i) optimize the individual expression of several genes and (ii) after one round of optimization in which one enzyme was clamped at the optimal level, continue the optimization of other enzymes in the pathway. With the systems available until now, one would then quickly run out of expression systems to use, but with our method, one can in principle continue the optimization numerous times.

In this report, the method for generating synthetic promoters of different strengths was illustrated for use in the gram-positive bacterium *L. lactis*. However, there is no obvious reason why the approach should be limited to this organism, and the fact that the same promoter library was also functional in the gram-negative bacterium *E. coli* suggests that the approach may be universally applicable to prokaryotic organisms. An exciting question is then, can the approach be extended to work for modulating gene expression in eukaryotic cells? Such experiments are under way, and the results are quite encouraging.

#### ACKNOWLEDGMENTS

We are deeply indebted to Regina Schürmann for excellent technical assistance.

This work was funded by the Danish Centre for Advanced Food Studies.

#### REFERENCES

- Boe, L. Personal communication.
- Casabadan, M. J., and S. N. Cohen. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**:179–207.
- de Ruyter, P. G., O. P. Kuipers, and W. M. de Vos. 1996. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62**:3662–3667.
- de Vos, W. M., and G. Simons. 1994. Gene cloning and expression systems in lactococci, p. 52–105. In M. J. Gasson and W. M. de Vos (ed.), *Genetics and biotechnology of lactic acid bacteria*. Blackie Academic & Professional, Glasgow, United Kingdom.
- Gasson, M. J. 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. *J. Bacteriol.* **154**:1–9.
- Heinrich, R., and T. A. Rapoport. 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector-strength. *Eur. J. Biochem.* **42**:89–95.
- Holo, H., and I. F. Nes. 1989. High-frequency transformation, by electroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. *Appl. Environ. Microbiol.* **55**:3119–3123.
- Israelsen, H., S. M. Madsen, A. Vrang, E. B. Hansen, and E. Johansen. 1995. Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-*lacZ* integrants with the new promoter probe vector, pAK80. *Appl. Environ. Microbiol.* **61**:2540–2547.
- Jensen, P. R. 1997. International patent application PCT/DK97/00342.
- Jensen, P. R., H. V. Westerhoff, and O. Michelsen. 1993. The use of *lac*-type promoters in control analysis. *Eur. J. Biochem.* **211**:181–191.
- Jensen, P. R., H. V. Westerhoff, and O. Michelsen. 1993. Excess capacity of  $H^+$ -ATPase and inverse respiratory control in *Escherichia coli*. *EMBO J.* **12**:1277–1282.
- Kacser, H., and J. A. Burns. 1973. The control of flux. *Symp. Soc. Exp. Biol.* **27**:65–104.
- Madsen, P. L. 1996. Transcription of the lactococcal temperate phage TP901-1. Ph.D. thesis. Department of Biological Chemistry, University of Copenhagen, Copenhagen, Denmark.
- Makrides, S. C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**:512–538.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor, Cold Spring Harbor Laboratory, N.Y.
- Miller, J. H. 1972. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Niederberger, P., R. Prasad, G. Miozzari, and H. Kacser. 1992. A strategy for increasing an in vivo flux by genetic manipulations. *Biochem. J.* **287**:473–479.
- Nilsson, D., and E. Johansen. 1994. A conserved sequence in tRNA and rRNA promoters of *Lactococcus lactis*. *Biochim. Biophys. Acta* **1219**:141–144.
- Pérez-Martín, J., F. Rojo, and V. de Lorenzo. 1994. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol. Rev.* **58**:268–290.
- Ruijter, G. J. G., P. W. Postma, and K. van Dam. 1991. Control of glucose metabolism by enzyme II<sup>Glc</sup> of the phosphoenolpyruvate-dependent phosphotransferase system in *Escherichia coli*. *J. Bacteriol.* **173**:6184–6191.
- Schaaff, I., J. Heinisch, and F. K. Zimmermann. 1989. Overproduction of glycolytic enzymes in yeast. *Yeast* **5**:285–290.
- Schickor, P., W. Metzger, W. Werel, H. Lederer, and H. Heumann. 1990. Topography of intermediates in transcription initiation of *E. coli*. *EMBO J.* **9**:2215–2220.
- Schneider, K., and C. F. Beck. 1986. Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene* **42**:37–48.
- Snoep, J. L., L. P. Yomano, H. V. Westerhoff, and L. O. Ingram. 1995. Protein burden in *Zymomonas mobilis*: negative flux and growth control due to overproduction of glycolytic enzymes. *Microbiology* **141**:2329–2337.
- Walsh, K., and D. E. Koshland, Jr. 1985. Characterization of rate-controlling steps in vivo by use of an adjustable expression vector. *Proc. Natl. Acad. Sci. USA* **82**:3577–3581.

## Isolation and characterization of mutants of firefly luciferase which produce different colors of light

Naoki Kajiyama and Eiichi Nakano

Research and Development Division, Kikkoman Corporation, 399 Noda, Noda-city, Chiba 278, Japan

The luciferase cDNA from the 'Genji' firefly, *Luciola cruciata*, was mutated with hydroxylamine to isolate mutant luciferases. Some of the isolated mutant enzymes produced different colors of light, ranging from green to red. Five such mutants, producing green ( $\lambda_{\max} = 558$  nm), yellow-orange ( $\lambda_{\max} = 595$  nm), orange ( $\lambda_{\max} = 607$  nm) and red light ( $\lambda_{\max} = 609$  and 612 nm), were analyzed. The mutations were found to be single amino acid changes, from Val239 to Ile, Pro452 to Ser, Ser286 to Asn, Gly326 to Ser and His433 to Tyr respectively.

**Key words:** color mutant/firefly luciferase/random mutagenesis/wavelength of maximum intensity

### Introduction

Firefly luciferase catalyzes the production of light from luciferin in the presence of ATP,  $Mg^{2+}$  and molecular oxygen (DeLuca and McElroy, 1978). This enzyme efficiently converts chemical energy into light with a quantum yield of 0.88 (Seliger and McElroy, 1960). Due to its high sensitivity and extreme specificity for ATP, luciferase has been used for assay of ATP in various biological samples (Ludin, 1981).

The luciferase cDNA from the Japanese firefly, *Luciola cruciata* ('Genji-botaru' in Japanese), has been cloned and analyzed in our laboratory (Masuda *et al.*, 1989). The primary structure of this luciferase deduced from the nucleotide sequence was shown to consist of 548 amino acids, with a total molecular weight of 60 024. This luciferase catalyzes a reaction that produces yellow-green light ( $\lambda_{\max} = 562$  nm), which is the same as that emitted by the North American firefly (Seliger and McElroy, 1964). It has been shown that the colors of light emitted by fireflies vary among species from green to yellow ( $\lambda_{\max} = 543$ –582 nm) (Seliger and McElroy, 1964). Since the substrate (D-luciferin) is the same for all species, the differences in the color of the light must be due to variations in the structure of the enzymes (McElroy and Seliger, 1966). Recently, cDNAs of luciferase from the bioluminescent click beetle, *Pyrophorus plagiophthalmus*, were cloned and their nucleotide sequences determined (Wood *et al.*, 1989a). These cDNA clones code for luciferases of four different types, distinguished by the colors of their bioluminescence. The amino acid sequences of these luciferases are 95–99% identical, and less than two or three amino acid changes are needed for the spectral shift in the color (Wood *et al.*, 1989b,c).

In the course of mutagenesis studies of luciferase cDNA from *Luciola cruciata*, we found that some mutants emitted different colors of light. Sequence analysis of these mutants revealed that the mutations were single amino acid changes.

### Materials and methods

#### Plasmid, *Escherichia coli* strain and media

Plasmid pGLf37 was constructed from pGLf1 by Mr H. Tatsumi in our laboratory (Masuda *et al.*, 1989). *Escherichia coli* strain JM 101 (*SupE*, *thi*,  $\Delta(lac-pro)$ , [*F'**traD36*, *lacI*  $\Delta$ M15, *proAB*]) (Yanish-Perron *et al.*, 1985) was used for the expression of luciferase cDNA. The *E. coli* cells were grown in LB broth (1% Difco tryptone/0.5% yeast extract/0.5% sodium chloride), and 50  $\mu$ g/ml ampicillin was added when necessary.

#### Mutagenesis and screening of 'color' mutants

Plasmid pGLf37 containing Genji-firefly luciferase cDNA was treated, according to the methods of Kironde *et al.* (1989), with 0.8 M hydroxylamine/0.1 M sodium phosphate/1 mM ethylenediaminetetraacetic acid (EDTA), pH 6.0, for 2 h at 65°C (Figure 1). The mutagen-treated plasmid was precipitated with ethanol and redissolved in 10 mM Tris-HCl/1 mM EDTA, pH 8.0, followed by transformation into *E. coli* JM 101. After 12 h at 37°C, colonies on LB/ampicillin plates were transferred to nitrocellulose filters. The filters were soaked with 0.5 mM luciferin in 100 mM sodium citrate buffer, pH 5.0 (Wood and DeLuca, 1987), and the colors of bioluminescence emitted by the colonies were monitored.

#### Purification of luciferase

*Escherichia coli* JM 101 cells harboring the mutant plasmid were cultured in 3 ml of LB broth containing ampicillin at 37°C for 12 h. The cultures, 2 ml each, were inoculated into 100 ml of LB broth containing ampicillin. After growth at 37°C for 6 h, the cultures were harvested. *Escherichia coli* pellets were resuspended in lysis buffer (100 mM potassium phosphate, pH 7.8/2 mM EDTA/1 mg lysozyme per ml), incubated on ice for 15 min and then frozen on dry ice. The frozen pellets were allowed to thaw at 25°C and cleared by centrifugation.

The lysates of *E. coli* were fractionated with ammonium sulfate; the fraction precipitated between 0.3 and 0.6 saturation was saved. The precipitate was dissolved with 25 mM Tris-HCl buffer, pH 7.8/1 mM EDTA/10%-saturated ammonium sulfate, and then loaded on an Ultrogel Aca34 gel filtration column (LKB). The active fraction was applied to a hydroxyapatite column (Tosoh, Tokyo, Japan), followed by elution with a 10–100 mM sodium phosphate gradient.

#### DNA sequencing

Various restriction fragments derived from mutant luciferase cDNA were subcloned into pUC118 or pUC119, and sequenced using a DNA sequencer model 373A (Applied Biosystems).

### Results and discussion

#### Screening of 'color' mutants

As shown in Figure 1, pGLf37, which is a plasmid directing the synthesis of active luciferase in *E. coli* under control of the *trp* promoter, was treated with hydroxylamine solution for 2 h at

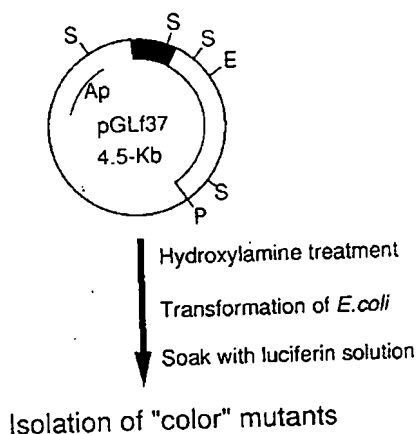


Fig. 1. Strategy of mutation. Plasmid pGLf37 containing Genji firefly luciferase cDNA was treated according to Materials and methods. Clear and solid portions show the luciferase cDNA and *trp* promoter respectively. Abbreviations: E, *EcoRV*; P, *Pst*I; S, *Ssp*I; Ap, ampicillin resistance gene.

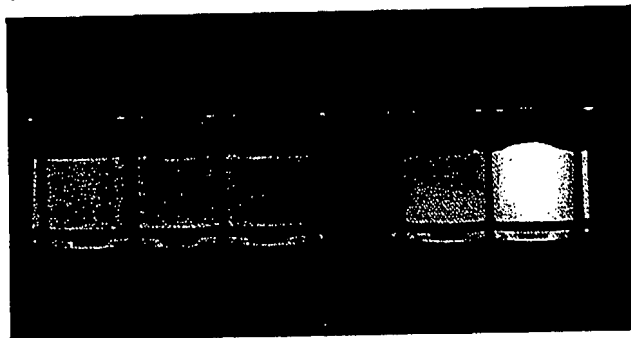


Fig. 2. The color of bioluminescence emitted by the wild-type and mutant luciferases. Left to right: yellow-green (wild-type luciferase); yellow-orange; orange; red; green; yellow. These luciferases were purified according to Materials and methods. Ten microliters of these luciferases were added to 400  $\mu$ l of substrate mix (25 mM glycylglycine, pH 7.8/5.4 mM  $MgSO_4$ /0.086 mM luciferin/2 mM ATP) to confirm the bioluminescence.

65°C. The plasmid was transformed into *E. coli* JM 101. To monitor the colors of bioluminescence, the transformants were soaked with luciferin solution. Subsequently, we isolated several mutants, producing colors of light varying from green to red.

#### Bioluminescence spectra of mutant luciferases

Mutant luciferases were purified to homogeneity as described in Materials and methods. Figure 2 shows the colors of light emitted by the purified enzymes. When their spectra were measured with a IMUC-7000 intensified multichannel photodetector (Otsuka Denshi, Osaka, Japan), the wavelengths of maximum intensity were 558 nm for green, 565 nm for yellow, 595 nm for yellow-orange, 607 nm for orange and 609 and 612 nm for red (Table I).

It is known that luciferase from the American firefly produces light with a peak intensity at ~560 nm (yellow-green) under optimal conditions (Seliger and McElroy, 1964). However, this peak can be affected by temperature, pH and metal ions. At low pH or in the presence of heavy metals, the emission peak is shifted toward the red, showing an emission of ~615 nm, but with a marked decrease in the quantum yield of the reaction (Seliger and McElroy, 1960, 1964). This phenomenon is also observed for Genji firefly luciferase. In the mutants C-M-1, 2, 3, 4 and 11, by contrast, the spectral peaks were shifted toward longer wavelengths under optimal conditions, with no detectable decrease of light intensity. Moreover, for mutants C-M-3 (red)

Table I. Wavelength of maximum intensity of light from wild-type and mutant luciferases

Luciferase	Color <sup>a</sup>	$\lambda_{max}$ (nm)	
		pH 7.8	pH 6.0
Genji	yellow-green	562	609
C-M-1	orange	607	614
C-M-2	red	609	611
C-M-3	red	612	612
C-M-4	yellow-orange	595	609
C-M-6	green	558	558
C-M-11	yellow	565	612

The spectra were measured with a IMUC-7000 intensified multichannel photodetector at pH 7.8 and pH 6.0. Except for pH, the condition for luminescence was the same as described in Figure 2.

<sup>a</sup>Color was confirmed at pH 7.8.

Table II. DNA sequence and amino acid sequence change in mutants

Mutant	Color	Base change	Amino acid change
C-M-1	orange	G857 → A	Ser286 → Asn
C-M-2	red	G976 → A	Gly326 → Ser
C-M-3	red	C1297 → T	His433 → Tyr
C-M-4	yellow-orange	C1354 → T	Pro452 → Ser
C-M-6	green	G715 → A	Val239 → Ile

and C-M-6 (green) there was no pH effect on emission spectra (Table I). For the other red (C-M-2), orange (C-M-1) and yellow-orange (C-M-4) mutants, the shift of spectral peak at pH 6.0 was detected slightly, but was not so large as for Genji luciferase. On the other hand, for the yellow mutant C-M-11, the shift was largest in the mutant luciferases and its reaction decay was similar to that for Genji luciferase. In contrast, the reaction decay in the C-M-3 and C-M-6 mutants at pH 6.0 was not as great as that of Genji wild-type luciferase (data not shown). It is uncertain whether the color shifts and decay rates of the light output at low pH are related to each other.

#### Amino acid sequences of mutants

Determination of the nucleotide sequences in some mutant luciferases (green, yellow-orange, orange and red) was carried out to identify the base changes present. As shown in Table II, the change in the yellow-orange mutant was found to be from CCA to TCA, resulting in a change of Pro to Ser at position 452. In the orange mutant, the alteration was from Ser to Asn at position 286. In the two red mutants, the changes for C-M-2 and C-M-3 were identified as Gly326 to Ser and His433 to Tyr respectively. The green mutant contained a change of Val to Ile at position 239. These results indicate that only a single amino acid substitution in a luciferase molecule is enough to produce the change in bioluminescence color.

Recently the nucleotide sequences of click beetle luciferase cDNAs were determined by Wood *et al.* (1989a). These are of four different types, distinguishable by the colors of light produced by the luciferases they code: green, yellow-green, yellow and orange. Fragments of the four different types were recombined to construct hybrid luciferases, and two groups of amino acids, each capable of producing a change in the spectrum of luciferase greater than 16 nm, were detected (Wood *et al.*, 1989b). For the first set, which contains the changes Arg223 to Glu and Leu238 to Val, the spectrum shifts from 560 to 577 nm. The spectrum for the other set, containing the changes Ser247 to Gly, Asp352 to Val and Ser358 to Thr, shifts from 560 to

580 nm. It is not known whether all the amino acids in each set are required for the spectral change.

In our experiments the colors of light were shown to be changed effectively by only one amino acid substitution (Table II). Four mutants showed an upward spectral shift of  $> 30$  nm, and the wavelength of their maximum light intensity was far longer than the orange of click beetle luciferase (593 nm). A red mutant (C-M-3), containing a change of His433 to Tyr, showed an especially large upward shift of 50 nm to 612 nm.

When the sequences of the mutant luciferases were compared with those of click beetle enzymes, no common amino acid sequence affecting the color of light was detected. Further, in the color mutants of firefly luciferase, there was no reversal of hydrophobicity or a large change in the conformational parameters of the secondary structure. For example, in the green mutant, the amino acid alteration was from Val to Ile. Both of these are hydrophobic, and conformational parameters for the  $\alpha$ -helix and  $\beta$ -sheet together with their charges are also similar (Chou and Fasman, 1978). When we estimated the secondary structures of wild-type and mutant luciferases from each of the primary structures using the algorithm of Garnier *et al.* (1978), no drastic structural changes were detected. Thus, it may be concluded that the different colors of bioluminescence were caused by only subtle differences in the tertiary structure of the luciferase molecule. It would be of interest to discover whether the color change, which is a unique parameter for luciferase, is related to the other enzymatic properties. Further studies, including examination of the properties of these color mutant enzymes and analysis of their tertiary structures, would elucidate the relationship between the structure and function of luciferase.

Firefly luciferase has been used for assay of ATP in various biological samples (Ludin, 1981). The mutant luciferases described above could be used more effectively for determining the amount of ATP in colored samples. In the case of red-colored samples, determination was found to be twice as sensitive using the mutant enzyme producing red light than with wild-type luciferase (data not shown).

Hydroxylamine treatment used in this study was a very simple and efficient method for introducing random base substitutions into luciferase cDNA. However, since this chemical mutagen causes only GC to AT transition mutations, mutants with limited substitution of amino acids can be obtained. To overcome this problem, Myers *et al.* (1985) used the method of treating single-stranded DNA with nitrous acid, formic acid and hydrazine, followed by the synthesis of the complementary strand with reverse transcriptase. Using this method for luciferase, various color mutants which cannot be found at present may be obtained.

In the present study, we succeeded for the first time in isolating several 'color' mutant luciferases. Sequence analysis revealed that only a single amino acid change in the 548 amino acids of luciferase is enough for the color variation. Analysis of other mutant luciferases is now in progress.

### Acknowledgements

We wish to thank Mr Hiroki Tatsumi for pGLF37 and Mrs Kazuko Saitoh for her technical assistance. We also wish to thank Mrs Hideko Otake for critical reading of the manuscript.

### References

- Chow, P.Y. and Fasman, G.D. (1978) In Meister, A. (ed.), *Advances in Enzymology*. Wiley, New York, Vol. 47, pp. 45–148.
- DeLuca, M. and McElroy, W.D. (1978) In DeLuca, M. (ed.), *Methods in Enzymology*. Academic Press, New York, Vol. 57, pp. 3–15.
- Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.*, 120, 97–120.
- Kironde, Fred A.S., Parsonage, D. and Senior, A.E. (1989) *J. Biochem.*, 259, 421–426.
- Ludin, A. (1981) In DeLuca, M. and McElroy, W.D. (eds), *Bioluminescence and Chemiluminescence*. Academic Press, New York, pp. 187–196.
- Masuda, T., Tatsumi, H. and Nakano, E. (1989) *Gene*, 77, 265–270.
- McElroy, W.D. and Seliger, H.H. (1966) In Hayashi, T. and Szent-Gyorgyi, A. (eds), *Molecular Architecture in Cell Physiology*. Prentice Hall, Englewood Cliffs, NJ, pp. 63–80.
- Myers, R.M., Lerman, L.S. and Maniatis, T. (1985) *Science*, 229, 242–247.
- Seliger, H.H. and McElroy, W.D. (1960) *Arch. Biochem. Biophys.*, 88, 136.
- Seliger, H.H. and McElroy, W.D. (1964) *Proc. Natl Acad. Sci. USA*, 52, 75–81.
- Wood, K.V., Lam, Y.A., Seliger, H.H. and McElroy, W.D. (1989a) *Science*, 244, 700–702.
- Wood, K.V., Lam, Y.A., McElroy, W.D. and Seliger, H.H. (1989b) *J. Biolumin. Chemilumin.*, 4, 31–39.
- Wood, K.V., Lam, Y.A. and McElroy, W.D. (1989c) *J. Biolumin. Chemilumin.*, 4, 289–301.
- Wood, K.V. and DeLuca, M. (1987) *Anal. Biochem.*, 161, 501–507.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, 33, 103.

## Isolation and expression of a cDNA encoding *Renilla reniformis* luciferase

(bioluminescence/*Renilla* luciferase/green fluorescent protein/gene expression)

W. WALTER LORENZ\*, RICHARD O. MCCANN\*, MATHEW LONGIARU†, AND MILTON J. CORMIER\*‡

\*Department of Biochemistry, University of Georgia, Athens, GA 30602; and †Hoffman-La Roche, Nutley, NJ 07110

Communicated by W. D. McElroy, December 26, 1990 (received for review, October 12, 1990)

**ABSTRACT** *Renilla reniformis* is an anthozoan coelenterate capable of exhibiting bioluminescence. Bioluminescence in *Renilla* results from the oxidation of coelenterate luciferin (coelenterazine) by luciferase [*Renilla*-luciferin:oxygen 2-oxidoreductase (decarboxylating), EC 1.13.12.5]. *In vivo*, the excited state luciferin-luciferase complex undergoes the process of nonradiative energy transfer to an accessory protein, green fluorescent protein, which results in green bioluminescence. *In vitro*, *Renilla* luciferase emits blue light in the absence of any green fluorescent protein. A *Renilla* cDNA library has been constructed in  $\lambda$ gt11 and screened by plaque hybridization with two oligonucleotide probes. We report here the isolation and characterization of a luciferase cDNA and its gene product. The recombinant luciferase expressed in *Escherichia coli* is identical to native luciferase as determined by SDS/PAGE, immunoblot analysis, and bioluminescence emission characteristics.

*Renilla reniformis* (class Anthozoa) is a bioluminescent soft coral found in shallow coastal waters of North America, which displays blue-green bioluminescence upon mechanical stimulation (1, 2). The components involved in *Renilla* bioluminescence have been described in detail (3). *Renilla* luciferase [*Renilla*-luciferin:oxygen 2-oxidoreductase (decarboxylating), EC 1.13.12.5] catalyzes the oxidative decarboxylation of coelenterazine in the presence of dissolved oxygen to yield oxyluciferin, CO<sub>2</sub>, and blue light ( $\lambda_{\text{max}}$  = 480 nm) (4). This reaction has a bioluminescence quantum yield of  $\approx 7\%$ . The stoichiometry of this reaction and the detailed mechanism leading to excited-state formation have been described (4, 5).

The color of *in vitro*-catalyzed bioluminescence changes from blue to green upon addition of submicromolar amounts of an energy-transfer acceptor green fluorescent protein (GFP), which has been purified from *Renilla* and characterized (6). This green fluorescence ( $\lambda_{\text{max}}$  = 509 nm) is identical to the *in vivo* emission in *Renilla*. The energy-transfer process is nonradiative; an increase in both the quantum yield (6) and calculated lifetimes has been determined for this process (7). Luciferase and GFP form a specific 1:1 rapid equilibrium complex in solution (8).

The elucidation of mechanisms involved in nonradiative energy transfer processes as well as determination of detailed structural information on both luciferase and GFP have been hindered by a lack of material. To overcome this, we have cloned, sequenced, and expressed in *Escherichia coli* a cDNA encoding *Renilla* luciferase.<sup>§</sup>

### MATERIALS AND METHODS

**Amino Acid Sequence Determination of *Renilla* Luciferase.** Native *Renilla* luciferase was isolated as described (4). Purified luciferase was digested with *Staphylococcal* protease

V-8 (9). The resulting peptides were purified by HPLC and subjected to NH<sub>2</sub>-terminal Edman sequencing as described (10). Based on these peptide sequences two 17-base oligonucleotide probes were synthesized with an Applied Biosystems DNA synthesizer at the Molecular Genetics Instrumentation Facility at the University of Georgia.

**Construction of a cDNA Library in  $\lambda$ gt11.** Live *R. reniformis* were collected at the University of Georgia Marine Institute located at Sapelo Island. The animals were frozen immediately in liquid N<sub>2</sub> and stored at  $-80^{\circ}\text{C}$ . Frozen tissue was ground to a fine powder in liquid N<sub>2</sub> with mortar and pestle. Total RNA was isolated from the frozen powder by the guanidine thiocyanate method (11), and poly(A)<sup>+</sup> RNA was isolated by oligo(dT)-cellulose chromatography (12). cDNA was synthesized by the method of Gubler and Hoffman (13). Phosphorylated *Eco*RI linkers (Collaborative Research) were ligated to the cDNAs, which were then digested with *Eco*RI. Separation of cDNA from free linkers after *Eco*RI digestion as well as size selection of cDNAs were accomplished by electrophoresis in low-melting-temperature agarose (NuSieve, FMC) (14). cDNAs were ligated into the *Eco*RI site of  $\lambda$ gt11 (15). The library was amplified in Y1088 cells (16) by a plate method (17).

**Isolation and DNA Sequence Determination of a Luciferase cDNA.** Oligonucleotide probes were 5' end-labeled with T4 polynucleotide kinase (Bethesda Research Laboratories) and [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol; 1 Ci = 37 GBq; ICN) to specific activities  $\geq 1 \times 10^8$  cpm/ $\mu\text{g}$  (18). A total of  $6 \times 10^5$  recombinant plaque-forming units were screened by plaque hybridization (19). Phage DNA was isolated as described (20). A luciferase cDNA, isolated from the clone  $\lambda$ RLuc-6, was subcloned into the M13 sequencing vectors mp18 and mp19, and sequencing templates were prepared (21). The DNA sequence of both strands was determined by the dideoxynucleotide chain-termination technique by using a Sequenase kit (United States Biochemical) and [ $\alpha$ -<sup>35</sup>S]dATP (400 Ci/mmol; Amersham) (22). The M13 universal primer and a  $\lambda$ gt11 sequencing primer (Amersham) were used to prime the sequencing reactions.

**Expression of Recombinant Luciferase (r-luciferase).** Positive clones were converted to lysogens in *E. coli* Y1089 cells (16). Lysogens were grown at permissive temperatures and induced with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG). Crude cell extracts were prepared and assayed for luciferase activity as described below. The plasmid pTZRLuc-1 was constructed by ligation of a 2.2-kilobase-pair (kbp) *Eco*RI/*Sst* I  $\lambda$ RLuc-6 restriction fragment into the plasmid pTZ18R (Pharmacia), which contains the *lacZ'* gene. *E. coli* TG-1 cells (23) were transformed with pTZRLuc-1 (24). Single colonies were isolated and grown at 37°C in LB

Abbreviations: GFP, green fluorescent protein; IPTG, isopropyl  $\beta$ -D-thiogalactopyranoside; r-luciferase, recombinant luciferase; ORF, open reading frame.

<sup>†</sup>To whom reprint requests should be addressed.

<sup>§</sup>The sequence reported in this paper has been deposited in the GenBank data base (accession no. M63501).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

medium containing ampicillin (100  $\mu$ g/ml) to an  $OD_{600}$  = 0.6–0.8 unit and induced with 1 mM IPTG for 4 hr. The cells were centrifuged at  $10,000 \times g$  and frozen solid at  $-20^{\circ}\text{C}$ . The pellets were thawed and resuspended 1:5 in 10 mM EDTA, pH 8, and lysozyme at 4 mg/ml (Sigma). After 20-min incubation at  $25^{\circ}\text{C}$ , the cells were placed on ice for 1 hr and then sonicated for 30 sec with a Branson cell disrupter. The cell lysate was clarified by centrifugation at  $30,000 \times g$ . The clarified lysate was used in subsequent bioluminescence assays and emission studies.

**Assay for *Renilla* Luciferase Activity and Determination of Emission Spectra.** Bioluminescence assays (4) were done with a Turner model TD-20e luminometer, and peak light intensities were determined. Bioluminescence intensity was converted to quanta per second by calibrating the instrument relative to a radioactive  $^{63}\text{Ni}$  light standard that emits in the 460- to 480-nm region (25). Corrected emission spectra were collected on an on-line computerized fluorimeter (26). A 100- $\mu$ l sample of a clarified pTZRLuc-1 cell extract was added to 1 ml of luciferase assay buffer (4) or to 1 ml of "energy-transfer buffer" containing  $1 \times 10^{-6}$  M GFP (8). An excess of coelenterazine (0.47 mM) dissolved in MeOH was added to maintain a strong emission signal.

**Genomic Southern Blot Analyses.** A 790-bp *EcoRI*/*Bam*HI cDNA restriction fragment was labeled to specific activities  $\geq 1 \times 10^9$  cpm/ $\mu$ g with both [ $\alpha$ - $^{32}\text{P}$ ]dATP and dCTP (4000 Ci/mmol, ICN) by the random hexamer-priming method (27). Genomic DNA was isolated from *Renilla* by a guanidine thiocyanate method developed for coelenterate DNA isolation (D. Prasher, personal communication). DNA samples were digested with the appropriate enzymes and resolved in a 0.8% agarose gel, followed by transfer to nitrocellulose filters (Schleicher & Schuell) (28). Aqueous hybridizations and washes were done at high stringencies as described for a homologous probe (17).

**Electrophoretic Analysis of Protein.** Protein samples were analyzed on 12.5% SDS/PAGE gels that were fixed and stained with Coomassie blue as described (29). Immunoblots were done as described (30). Proteins were transferred to nitrocellulose (Schleicher & Schuell) and incubated in a 1:50 dilution of rabbit anti-native luciferase antibody. Detection of the secondary antibody (horseradish peroxidase-conjugated goat anti-rabbit IgG) signal was determined according to the vendor's instructions (Bio-Rad).

**Computer-Facilitated DNA and Amino Acid Sequence Analyses.** The DNA sequence was compiled and manipulated using MicroGenie sequence software (Beckman).

## RESULTS

**Synthesis of Luciferase Oligonucleotides.** Seven luciferase peptides (V8-1–V8-7) were purified by HPLC, and their amino acid sequences were determined. Two of the peptides contained regions of relatively low codon degeneracy. Amino acid sequence from these regions was used to synthesize two mixed-sequence 17-base oligonucleotide hybridization probes with the following sequences; RLP-1 {GAR-AAYAAAY-TTY-TTY-GT} and RLP-2 {AAR-AAR-TTY-CCN-AAY AT}, which are 32- and 64-fold degenerate, respectively.

**Nucleotide and Deduced Amino Acid Sequence Analyses of *Renilla* Luciferase.** Six clones were isolated from the *Renilla* cDNA library. One clone,  $\lambda$ RLuc-6, hybridized to both oligonucleotide probes. The cDNA insert could not be isolated after *EcoRI* digestion, as one of the linker sites was lost during cloning. A double digest with *EcoRI* and *Sst* I produced a 2.2-kbp fragment that contained a 1.2-kbp cDNA with 1 kbp of  $\lambda$ gt11 DNA at the 3' end. This fragment was subcloned into the M13 vectors mp18 and mp19. DNA sequencing provided the locations of six base restriction sites (Fig. 1), which were used to generate specific sequencing subclones. The entire 1.2-kbp luciferase cDNA was sequenced on both strands.

The cDNA, excluding the *EcoRI* linkers, is 1196 nucleotides long and encodes an open reading frame (ORF) of 314 amino acids (Fig. 2). Although an ATG in-frame codon is found at the 5' end of the cDNA, the intrinsic mRNA may contain additional 5' coding nucleotides. If the first ATG codon in the ORF is designated as the initiation codon, the predicted 311 amino acid sequence is essentially identical in size (34 kDa) and composition to native *Renilla* luciferase (4).

Comparison of the deduced amino acid sequence with the native peptides reveals that  $\lambda$ RLuc-6 encodes a luciferase cDNA (Fig. 2). One discrepancy lies at amino acid residue 222, which is leucine in the peptide sequence but tryptophan in the deduced sequence. Sequencing autoradiograms from this region of the clone have been examined carefully and found free of any irregularities. The protein sequence also contains a consensus N-linked glycosylation site (Asn-Xaa-Ser) beginning at residue 92.

**Genomic Southern Analysis.** A *Renilla* genomic Southern blot was probed with a 790-base-pair (bp) *EcoRI*/*Bam*HI luciferase cDNA restriction fragment (Fig. 3). The *Bam*HI digest, lane A, contains two hybridizing bands as does the *Sma* I digest, lane B. The *Bgl* II digest, lane C, contains four bands. If luciferase is encoded by a single gene containing no introns, a single band would be expected in the *Bam*HI and *Sma* I digests, as these two sites are not spanned by the hybridization probe. Similarly, two bands would be expected in the *Bgl* II digest. That the *Bam*HI and *Sma* I digests contain two hybridizing bands shows either that there is more than one luciferase gene or that the luciferase gene(s) has introns containing *Bam*HI and *Sma* I sites. The four bands seen in the *Bgl* II could be explained by two very large introns containing *Bgl* II sites. When genomic DNA was digested with restriction enzymes having no sites within the cDNA sequence, there were always at least two or more hybridizing bands (data not shown). These results suggest that luciferase is encoded for by more than one gene, which may or may not contain introns.

**Luciferase Expression in *E. coli*.** The  $\lambda$ RLuc-6 lysogen is capable of low-level r-luciferase expression as determined by light emission from clarified, crude extracts ( $5 \times 10^{10}$  hv $\cdot$ sec $^{-1}$  $\cdot$ ml $^{-1}$ ). When these cells are induced with 1 mM IPTG, light emission decreases by 2-fold; this happens because the cDNA is reversely oriented with respect to the  $\lambda$ gt11 *lacZ* promoter. Presumably, when IPTG is absent, the luciferase gene is transcribed from a promoter in the right end of  $\lambda$ gt11, as reported (31).

The 2.2-kbp *EcoRI*/*Sst* I fragment was subcloned into the plasmid pTZ18R, which uses the *lacZ* promoter. The ORF of



FIG. 1. Location of six base restriction enzyme sites within the luciferase cDNA. The boxed region defines the ORF. The *EcoRI* site at the 5' end is a synthetic linker site.

1	AGC	TTA	AAG	ATG	ACT	TCG	AAA	GTT	TAT	GAT	CCA	GAA	CAA	AGG	AAA	CGG	ATG	ATA	ACT	GGT	60
1	Ser	Leu	Lys	Met	Thr	Ser	Lys	Val	Tyr	Asp	Pro	Glu	Gln	Arg	Lys	Arg	Met	Ile	Thr	Gly	20
61	CCG	CAG	TGG	TGG	GCC	AGA	TGT	AAA	CAA	ATG	AAT	GTT	CTT	GAT	TCA	TTT	ATT	AAT	TAT	TAT	120
21	Pro	Gln	Trp	Trp	Ala	Arg	Cys	Lys	Gln	Met	Asn	Val	Leu	Asp	Ser	Phe	Ile	Asn	Tyr	Tyr	40
121	GAT	TCA	GAA	AAA	CAT	GCA	GAA	AAT	GCT	GTT	ATT	TTT	TTA	CAT	GGT	AAC	GCG	GCC	TCT	TCT	180
41	Asp	Ser	Glu	Lys	His	Ala	Glu	Asn	Ala	Val	Ile	Phe	Leu	His	Gly	Asn	Ala	Ala	Ser	Ser	60
181	TAT	TTA	TGG	CGA	CAT	GTT	GTG	CCA	CAT	ATT	GAG	CCA	GTA	GCG	CGG	TGT	ATT	ATA	CCA	GAT	240
61	Tyr	Leu	Trp	Arg	His	Val	Val	Pro	His	Ile	Glu	Pro	Val	Ala	Arg	Cys	Ile	Ile	Pro	Asp	80
241	CTT	ATT	GGT	ATG	GCC	AAA	TCA	GCG	AAA	TCT	GGT	AAT	GGT	TCT	TAT	AGG	TTA	CTT	GAT	CAT	300
81	Leu	Ile	Gly	Met	Gly	Lys	Ser	Gly	Lys	Ser	Gly	Asn	Gly	Ser	Tyr	Arg	Leu	Leu	Asp	His	100
301	TAC	AAA	TAT	CTT	ACT	GCA	TGG	TTT	GAA	CTT	CTT	AAT	TTA	CCA	AAG	AAG	ATC	ATT	TTT	GTC	360
101	Tyr	Lys	Tyr	Leu	Thr	Ala	Trp	Phe	Glu	Leu	Leu	Asn	Leu	Pro	Lys	Lys	Ile	Ile	Phe	Val	120
361	GGC	CAT	GAT	TGG	GGT	GCT	TGT	TTG	GCA	TTT	CAT	TAT	AGC	TAT	GAG	CAT	CAA	GAT	AAG	ATC	420
121	Gly	His	Asp	Trp	Gly	Ala	Cys	Leu	Ala	Phe	His	Tyr	Ser	Tyr	Glu	His	Gln	Asp	Lys	Ile	140
421	AAA	GCA	ATA	GTT	CAC	GCT	GAA	AGT	GTA	GTA	GAT	GTG	ATT	GAA	TCA	TGG	GAT	GAA	TGG	CCT	480
141	Lys	Ala	Ile	Val	His	Ala	Glu	Ser	Val	Val	Asp	Val	Ile	Glu	Ser	Trp	Asp	Glu	Trp	Pro	160
481	GAT	ATT	GAA	GAA	GAT	ATT	GCG	TTG	ATC	AAA	TCT	GAA	GAA	GGA	GAA	AAA	ATG	GTT	TTG	GAG	540
161	Asp	Ile	Glu	Glu	Asp	Ile	Ala	Leu	Ile	Lys	Ser	Glu	Glu	Gly	Glu	Lys	Met	Val	Leu	Glu	180
541	AAT	AAC	TTC	TTC	GTG	GAA	ACC	ATG	TTG	CCA	TCA	AAA	ATC	ATG	AGA	AAG	TTA	GAA	CCA	GAA	600
181	Asn	Asn	Phe	Phe	Val	Glu	Thr	Met	Leu	Pro	Ser	Lys	Ile	Met	Arg	Lys	Leu	Glu	Pro	Glu	200
601	GAA	TTT	GCA	GCA	TAT	CTT	GAA	CCA	TTC	AAA	GAG	AAA	GGT	GAA	GTT	CGT	CGT	CCA	ACA	TTA	660
201	Glu	Phe	Ala	Ala	Tyr	Leu	Glu	Pro	Phe	Lys	Glu	Lys	Gly	Glu	Val	Arg	Arg	Pro	Thr	Leu	220
661	TCA	TGG	CCT	CGT	GAA	ATC	COG	TTA	GTA	AAA	GGT	GGT	AAA	CCT	GAC	GTT	GTA	CAA	ATT	GTT	720
221	Ser	Trp	Pro	Arg	Glu	Ile	Pro	Leu	Val	Lys	Gly	Gly	Lys	Pro	Asp	Val	Val	Gln	Ile	Val	240
721	AGG	AAT	TAT	AAT	GCT	TAT	CTA	CGT	GCA	AGT	GAT	GAT	TTA	CCA	AAA	ATG	TTT	ATT	GAA	TCG	780
241	Arg	Asn	Tyr	Asn	Ala	Tyr	Leu	Arg	Ala	Ser	Asp	Asp	Leu	Pro	Lys	Met	Phe	Ile	Glu	Ser	260
781	GAT	CCA	GGA	TTC	TTT	TCC	AAT	GCT	ATT	GTT	GAA	GCG	GCC	AAG	AAG	TTT	CCT	AAT	ACT	GAA	840
261	Asp	Pro	Gly	Phe	Phe	Ser	Asn	Ala	Ile	Val	Glu	Gly	Ala	Lys	Lys	Phe	Pro	Asn	Thr	Glu	280
841	TTT	GTC	AAA	GTA	AAA	GGT	CTT	CAT	TTT	TCG	CAA	GAA	GAT	GCA	CCT	GAT	GAA	ATG	GGA	AAA	900
281	Phe	Val	Lys	Val	Lys	Gly	Leu	His	Phe	Ser	Gln	Glu	Asp	Ala	Pro	Asp	Glu	Met	Gly	Lys	300
901	TAT	ATC	AAA	TCG	TTC	GTT	GAG	CGA	GTT	CTC	AAA	AAT	GAA	CAA	TAA	TACTTT	GGTTTTTTAT				960
301	Tyr	Ile	Lys	Ser	Phe	Val	Glu	Arg	Val	Leu	Lys	Asn	Glu	Gln							314
963	TTACATTTT	CCCGGGTTA	ATAATATAA	TGTCATTTT	AACAATTTA	TTTAACTGA	ATATTCACA														1032
1033	GGGAACATTC	ATATATGTTG	ATTAATTAG	CTCGAATTT	ACTCTGTCAT	ATCATTGTTG	AATATTACCT														1102
1103	CTTCAATGA	AACTTTATAA	ACAGTGGTTC	AATTAATTAA	TATATATTAT	AATTACATT	GITATGTAAT														1172
1173	AAACTCGTT	TTATTATAA	AAAA																		1196

FIG. 2. Nucleotide sequence and translated amino acid sequence of the *Renilla* luciferase cDNA. Putative and known translation control elements, as well as oligonucleotide hybridization sites, are underlined. Positions of native luciferase peptide sequences are boxed and, except at one residue (+), are identical to the deduced amino acid sequence obtained from the luciferase cDNA. Some of the native luciferase peptide sequences overlap at glutamic acid residues.

the cDNA is not in frame with the *lacZ'* gene ORF of pTZ18R. Supernatants were prepared from IPTG-induced pTZRLuc-1 cells, as described, and the level of luciferase expression was measured by the standard luciferase assay. A high level of r-luciferase activity,  $2 \times 10^{15}$  hv·sec<sup>-1</sup>·ml<sup>-1</sup>, is detected in clarified crude extracts of pTZRLuc-1 cells. This level of activity is 7-fold greater than in uninduced pTZRLuc-1 cells.

A prominent protein band ( $M_r = 34,000$ ) migrating to the position of native luciferase is seen after SDS/PAGE of pTZR-

Luc-1 crude extracts (Fig. 4). Crude extracts of IPTG-induced pTZRLuc-1 cells were analyzed by immunoblotting (Fig. 5). The protein band that reacts with antiluciferase antibody, lane B, corresponds to the same band seen in the Coomassie-stained gel (Fig. 4). Native luciferase was used as a positive control, lane A. No signal is detectable in the crude extract of pTZ18R cells, lane C. A duplicate filter incubated with preimmune serum showed no detectable signal.

**Bioluminescence Emission Spectra.** The r-luciferase-catalyzed bioluminescence emission spectrum (Fig. 6a) is



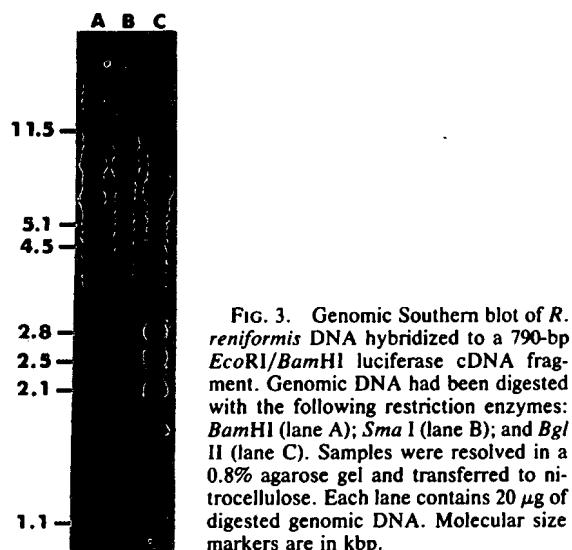


FIG. 3. Genomic Southern blot of *R. reniformis* DNA hybridized to a 790-bp *EcoRI/BamHI* luciferase cDNA fragment. Genomic DNA had been digested with the following restriction enzymes: *Bam*HI (lane A); *Sma*I (lane B); and *Bgl*II (lane C). Samples were resolved in a 0.8% agarose gel and transferred to nitrocellulose. Each lane contains 20  $\mu$ g of digested genomic DNA. Molecular size markers are in kbp.

very similar to that seen with native luciferase (32). The r-luciferase emission spectrum has a  $\lambda_{\max} = 480$  nm and a slight shoulder at 400 nm, which correspond to emission from the excited-state oxyluciferin monoanion and neutral species, respectively. Disproportionation between these species is sensitive to environmental factors (7); thus, this spectrum indicates the strong similarity of the active-site environment between r-luciferase and the native enzyme. Although an increase in quantum yield has yet to be determined, r-luciferase can clearly transfer energy in the presence of *Renilla* GFP (Fig. 6b). The emission spectrum dramatically shifted

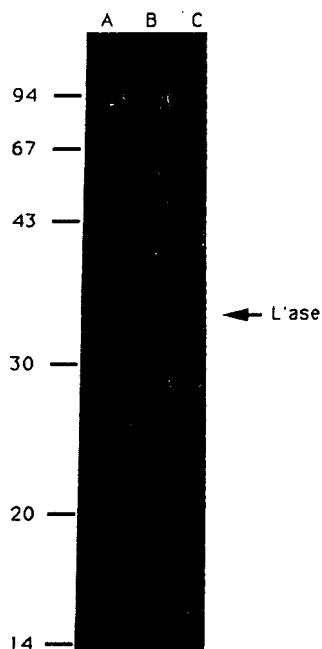


FIG. 4. SDS/PAGE analysis of total protein from IPTG-induced *E. coli* cells transformed with either pTZRLuc-1 or pTZ18R. Ten-milliliter cultures were grown to an  $OD_{600} = 0.8$  and induced with 1 mM IPTG for 4 hr. One-milliliter of cell culture,  $OD_{600} = 5.0$ , was pelleted and resuspended in 0.5 ml of SDS sample buffer. Samples were boiled for 5 min, and 20  $\mu$ l was loaded per lane: native luciferase (10  $\mu$ g) (lane A); pTZRLuc-1 cells (lane B); and pTZ18R cells (lane C). Molecular weight ( $M_r \times 10^{-3}$ ) standard positions are indicated. Arrow shows position of native luciferase (L'ase).

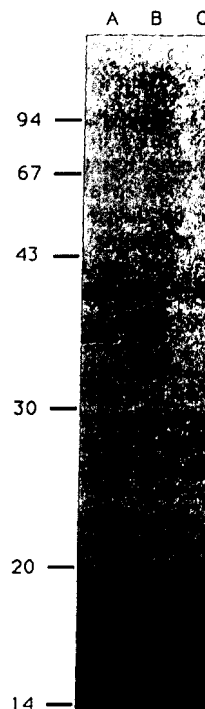


FIG. 5. Immunoblot analysis of total protein. Sample preparation and electrophoresis were the same as in Fig. 4. Native luciferase (2  $\mu$ g) (lane A); 10  $\mu$ l of pTZRLuc-1 cell extract (lane B); and 10  $\mu$ l of pTZ18R cell extract (lane C). Molecular weight ( $M_r \times 10^{-3}$ ) standard positions are indicated.

from the broad emission band generated by r-luciferase to the narrow, structured emission band ( $\lambda_{\max} = 509$ ) seen when GFP is present. The emission spectrum generated with r-luciferase and GFP is very similar to the same spectrum generated with native luciferase and GFP (33).

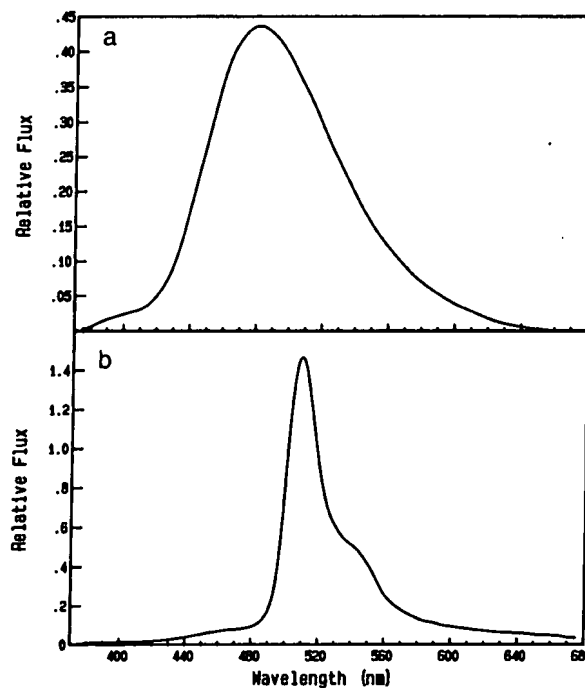


FIG. 6. Bioluminescence emission spectra generated with crude r-luciferase and r-luciferase plus GFP. Crude pTZRLuc-1 cell extracts were prepared, as described, and 100  $\mu$ l,  $\approx 1 \times 10^{-6}$  M r-luciferase, as determined by peak light emission, was used to generate each spectrum. (a) Emission spectrum of crude r-luciferase. (b) Spectrum that results when  $1 \times 10^{-6}$  M *Renilla* GFP is added to crude r-luciferase in energy-transfer buffer.

## DISCUSSION

This work describes the isolation of a 1.2-kbp *R. reniformis* luciferase cDNA capable of directing the expression of r-luciferase. The cDNA contains an ORF encoding a 314-amino acid sequence in which all of the native luciferase peptide sequences obtained from V8-protease digestion are found. Rescreening the cDNA library with a 790-bp luciferase cDNA fragment as a hybridization probe has failed to produce other clones that contain the 5' noncoding region; therefore, whether the luciferase cDNA is full-length is not known. This uncertainty can be resolved by sequencing genomic clones corresponding to the 5' end of the luciferase gene.

The genomic Southern hybridization data indicates that *Renilla* luciferase is probably encoded by more than one gene, which may or may not contain intervening sequences. Further characterization of luciferase genomic clones will be required before the genetic organization of the luciferase gene(s) can be defined.

A putative initiation codon located at triplet position 4 of the ORF may be the translation initiation site for *Renilla* luciferase; the 311-amino acid sequence is essentially identical to native luciferase with respect to its composition and predicted molecular weight. Irrespective of whether this cDNA is full length, the luciferase that it encodes is expressed in pTZRLuc-1 cells and is catalytically active. The expression data demonstrate that r-luciferase is the same size as native luciferase on SDS/PAGE gels and is reactive with polyclonal rabbit antibodies raised against native *Renilla* luciferase. Expression of r-luciferase from the plasmid pTZRLuc-1 is "leaky" because activity can be detected from uninduced cell cultures. The luciferase cDNA ORF is not in frame with the short *lacZ'* ORF contained in this construct. Any translation product initiating at the  $\beta$ -galactosidase sequence of pTZRLuc-1 would be terminated at a stop codon immediately adjacent to the putative initiation codon in the luciferase cDNA. Thus, the r-luciferase seen in SDS/PAGE gels does not contain any  $\beta$ -galactosidase sequence. We propose that expression of r-luciferase by pTZRLuc-1 is due to a translation coupling mechanism (34).

r-luciferase displays two very important characteristics of native luciferase: the ability to catalyze coelenterazine oxidation with the concomitant emission of blue light ( $\lambda_{\max} = 480$  nm) and the ability to transfer energy to *Renilla* GFP with the production of green light ( $\lambda_{\max} = 509$  nm). The two emission bands at 400 nm and 480 nm in the r-luciferase spectrum verify the strong similarity between the native and recombinant proteins and suggest that the integrity of the luciferase active site has been maintained. Furthermore, that energy transfer occurs in the presence of GFP shows that the luciferase domain(s) required for the interaction between luciferase and GFP is present in r-luciferase. Once pure r-luciferase is available, energy transfer quantum yield measurements will offer a more quantitative determination of the efficiency of the nonradiative energy-transfer process. Finally, the data demonstrate that N-linked glycosylation is not required for luciferase activity because *E. coli* do not perform this modification (35). r-luciferase in *E. coli* crude extracts behaves like the native enzyme by every criterion examined thus far.

We thank Yu-Ching Pan and Jeffery Hulmes (Roche Diagnostic Systems) for the partial amino acid sequence data on native luciferase; John Wunderlich for the synthesis of oligonucleotide probes; Dennis O'Kane and John Wampler for help in obtaining bioluminescence emission spectral data; Leon Dure, Claiborne

Glover, and Douglas Prasher for their critical comments; and Judy Gray for technical assistance. This is contribution number 659 from the University of Georgia Marine Institute, Sapelo Island, GA. This work was funded, in part, by a grant from ELA Technologies Inc.

- Harvey, E. N. (1952) in *Bioluminescence*, ed. Harvey, E. N. (Academic, New York), pp. 148–194.
- Morin, J. G. (1974) in *Coelenterate Biology: Reviews and Perspectives*, eds. Muscatine, L. & Lenhoff, H. M. (Academic, New York), pp. 397–438.
- Cormier, M. J. (1978) in *Bioluminescence in Action*, ed. Herring, P. J. (Academic, New York), pp. 75–108.
- Matthews, J. C., Hori, K. & Cormier, M. J. (1977) *Biochemistry* 16, 85–91.
- Hart, R. C., Stempel, K. E., Boyer, P. D. & Cormier, M. J. (1978) *Biochem. Biophys. Res. Commun.* 81, 980–986.
- Ward, W. W. & Cormier, M. J. (1979) *J. Biol. Chem.* 254, 781–788.
- Hart, R. C., Matthews, J. C., Hori, K. & Cormier, M. J. (1979) *Biochemistry* 18, 2204–2210.
- Ward, W. W. & Cormier, M. J. (1978) *Photochem. Photobiol.* 27, 389–396.
- Houmar, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. USA* 69, 3506–3509.
- Charbonneau, H., Walsh, K. A., McCann, R. O., Prendergast, F. G., Cormier, M. J. & Vanaman, T. C. (1985) *Biochemistry* 24, 6762–6771.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 18, 5294–5299.
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408–1412.
- Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263–269.
- Weislander, L. (1979) *Anal. Biochem.* 98, 305–309.
- Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1194–1198.
- Ausubel, F. M., Brert, R., Kingston, R. E., Morre, D. D., Seidmon, J. G., Smith, J. A. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Chaconas, G. & van de Sande, J. H. (1980) *Methods Enzymol.* 68, 75–88.
- Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1585–1588.
- Grossberger, D. (1987) *Nucleic Acids Res.* 15, 6737.
- Yanisch-Perron, C., Viera, J. & Messing, J. (1985) *Gene* 33, 103–119.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Gay, N. J. & Walker, J. E. (1985) *Biochem. J.* 225, 707–712.
- Kushner, S. R. (1978) in *Genetic Engineering*, eds. Boyer, H. W. & Nicosia, S. (Elsevier Biomed., Amsterdam), pp. 17–23.
- O'Kane, D. J. & Lee, J. (1990) *Photochem. Photobiol.* 52, 723–734.
- Mulkerrin, M. G. & Wampler, J. E. (1982) *Anal. Chem.* 54, 1778–1782.
- Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* 132, 6–13.
- Southern, E. M. (1975) *J. Mol. Biol.* 98, 503–517.
- Laemmli, U. K. (1970) *Nature (London)* 227, 680–685.
- Burnett, W. H. (1981) *Anal. Biochem.* 112, 195–203.
- Chirala, S. S. (1986) *Nucleic Acids Res.* 14, 5935.
- Matthews, J. C., Hori, K. & Cormier, M. J. (1977) *Biochemistry* 16, 5217–5220.
- Ward, W. W. & Cormier, M. J. (1978) *Methods Enzymol.* 57, 257–267.
- Stormo, G. D. (1986) in *Maximizing Gene Expression*, eds. Rezinikoff, W. & Gold, L. (Butterworths, Boston), pp. 195–224.
- Kornfeld, R. & Kornfeld, S. (1985) *Annu. Rev. Biochem.* 54, 631–664.

## RESEARCH

# Directed evolution of the surface chemistry of the reporter enzyme $\beta$ -glucuronidase

Ichiro Matsumura<sup>1</sup>, John B. Wallingford<sup>2,3</sup>, Neeraj K. Surana<sup>1,4</sup>, Peter D. Vize<sup>2</sup>, and Andrew D. Ellington<sup>1\*</sup>

<sup>1</sup>Institute of Cellular and Molecular Biology, ICMB A4800/MBB 3.424, 26th and Speedway, University of Texas, Austin, TX 78712. <sup>2</sup>Center for Developmental Biology, Department of Zoology, University of Texas, Austin, TX 78712. <sup>3</sup>Present address: Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720. <sup>4</sup>Present address: Division of Biology and Biomedical Sciences, Washington University, St. Louis, MO 63108. \*Corresponding author (e-mail: andy.ellington@mail.utexas.edu).

Received 29 December 1999; accepted 4 May 1999

The use of the *Escherichia coli* enzyme  $\beta$ -glucuronidase (GUS) as a reporter in gene expression studies is limited due to loss of activity during tissue fixation by glutaraldehyde or formaldehyde. We have directed the evolution of a GUS variant that is significantly more resistant to both glutaraldehyde and formaldehyde than the wild-type enzyme. A variant with eight amino acid changes was isolated after three rounds of mutation, DNA shuffling, and screening. Surprisingly, although glutaraldehyde is known to modify and cross-link free amines, only one lysine residue was mutated. Instead, amino acid changes generally occurred near conserved lysines, implying that the surface chemistry of the enzyme was selected to either accept or avoid glutaraldehyde modifications that would normally have inhibited function. We have shown that the GUS variant can be used to trace cell lineages in *Xenopus* embryos under standard fixation conditions, allowing double staining when used in conjunction with other reporters.

Keywords:  $\beta$ -glucuronidase, reporter gene, in vitro evolution, directed evolution, DNA shuffling, *Xenopus laevis*

Since plants express endogenous  $\beta$ -galactosidase activity, *lacZ* cannot be employed as a reporter gene<sup>1</sup>. Instead, the *Escherichia coli*  $\beta$ -glucuronidase gene (*gusA*, formerly *uidA*) has been developed as a reporter gene for plants, and has been widely used for over a decade<sup>2</sup>. Both chromogenic and fluorogenic GUS substrates have been synthesized<sup>3</sup>, allowing rapid nonradioactive assays. The GUS enzyme is stable and active under a variety of conditions<sup>1</sup>, even when fused to other sequences<sup>4</sup>.

The utility of GUS as a reporter, however, has been constrained in three ways. First, many animal systems, and some plants and plant-associated bacteria express endogenous glucuronidase activities<sup>2,3</sup>. Second, GUS activity is greatly reduced during tissue fixation by glutaraldehyde or formaldehyde, making it necessary to trade off retention of activity for preservation of tissue structure<sup>5</sup>. Third, both of these considerations drastically restrict the use of GUS as a reporter gene in vertebrate systems<sup>6</sup>.

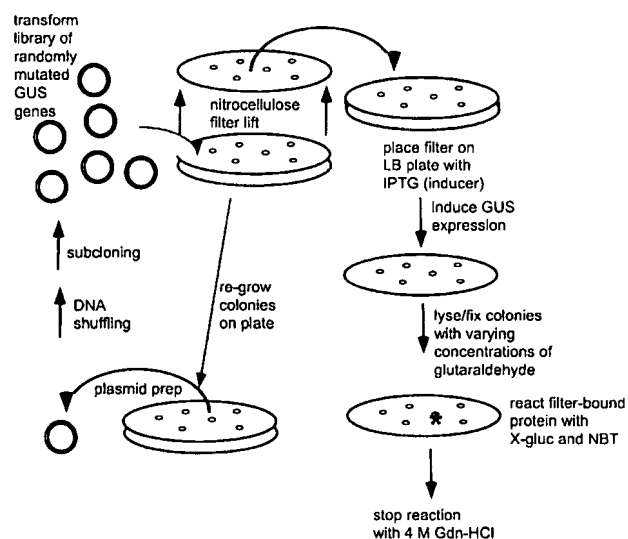
Enzymatic inactivation by aldehydes is largely due to the formation of Schiff bases with surface-accessible lysine residues<sup>7</sup>. While the removal of lysine residues by directed mutation might render an enzyme more resistant to fixatives, many surface lysines are critical for function and cannot be readily changed. The sequences of the *E. coli*<sup>8</sup>, human<sup>9</sup>, mouse<sup>10</sup>, rat<sup>11</sup>, and dog<sup>12</sup> homologs are known. Six of the 27 lysine residues in the *E. coli* protein are conserved in the other species and thus are likely essential. Moreover, to find what combination of the 27 lysine residues could be changed in order to increase resistance to fixatives without abrogating enzyme activity would require constructing and assaying a dauntingly large number of mutant enzymes. Therefore, in order to alter the surface chemistry of GUS, either to avoid or to accommodate aldehyde modifications without loss of enzyme activity, we employed a random mutation-

approach similar to those previously proven useful for altering enzyme substrate specificity<sup>13</sup> or thermostability<sup>14</sup>.

## Results

**Directed evolution of glutaraldehyde-resistant variants.** Random mutations were initially introduced into the *gusA* structural gene by mutagenic PCR<sup>15</sup>. Mutated PCR products were ligated into the expression vector *gusA*-pBSA and transformed into *E. coli*. When the library was induced on plates containing the chromogenic GUS substrate, 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronide (X-gluc), approximately 80% of the colonies were visibly less green than control colonies expressing only the chromosomal gene (see Experimental protocol).  $\beta$ -Glucuronidase functions as a tetramer<sup>16</sup>, so it was likely that many of the mutations in the highly expressed, plasmid-borne library had a dominant negative effect on the function of the chromosomal gene. This did not deter us from utilizing this library for screening experiments, since successive rounds of DNA shuffling should efficiently select against neutral or deleterious mutations<sup>13</sup>.

Nine thousand replica-plated colonies, each expressing a randomly mutated *gusA* gene, were exposed to buffer containing 0.2% glutaraldehyde for 20 min. The colony remnants were then incubated in buffer containing X-gluc and the histochemical indicator, nitroblue tetrazolium (NBT) (Fig. 1). The catalytic activity of the wild-type enzyme is greatly diminished under those conditions, indicating that the glutaraldehyde disrupts the cell membranes and covalently modifies many intracellular proteins, including GUS (cf. Figs 2A and B). Of all the variants examined, only 10 colonies reproducibly exhibited greater catalytic activity than control colonies expressing wild-type *gusA* (Fig. 2B). The corresponding colonies on master plates were isolated, and their expression vectors purified

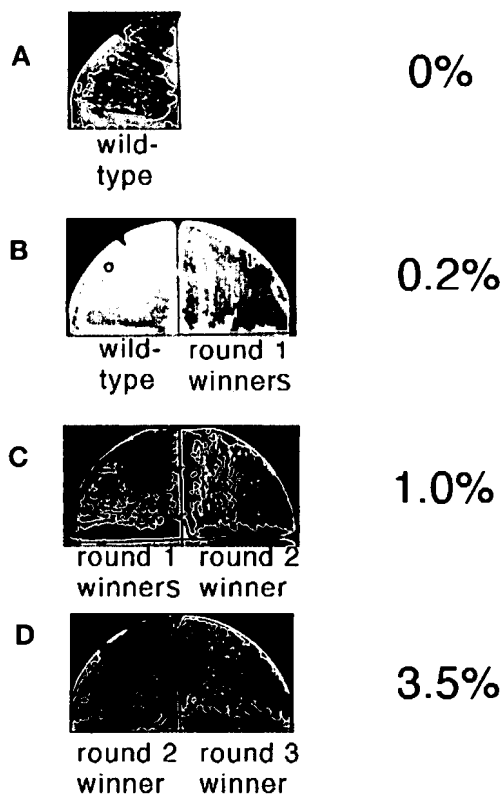


**Figure 1.** Screen for glutaraldehyde-resistant  $\beta$ -glucuronidase (GUS) function (sequence from top left). A library of randomly mutated  $\beta$ -glucuronidase genes (*gusA*) is subcloned into an inducible expression vector and transformed into *Escherichia coli*. The resulting colonies are transferred to a nitrocellulose filter, which is overlaid upon an agarose plate containing an inducer and incubated for 12–24 h at 37°C. The filter-bound colonies are incubated in buffer containing glutaraldehyde, then transferred to buffer containing the histochemical indicators of  $\beta$ -glucuronidase, X-gluc and NBT. The brief incubation in 4 M guanidine HCl (Gdn-HCl) arrests color development. Colonies that retain GUS activity are isolated from the original plate and randomly recombined by DNA shuffling for the next round of screening.

and pooled. The variant *gusA* genes were amplified using the PCR and randomly recombined by DNA shuffling<sup>17</sup>. We then screened 6,000 random recombinants in a second round for variants that retained catalytic activity after a 20 min incubation in 1.0% glutaraldehyde. Nine colonies contained variants that exhibited greater residual catalytic activity than the most resistant clone isolated in the first round of screening (Fig. 2C). These variants were again pooled, amplified, and randomly recombined. Then, 6,000 recombinants were screened in a third round for variants that retained catalytic activity after a 20 min incubation in 3.5% glutaraldehyde. Again, nine improved clones were isolated, one of which (*GUS*<sup>AR</sup>) reproducibly showed the greatest activity under the most stringent conditions (Fig. 2D).

**In vitro characterization of *GUS*<sup>AR</sup>.** To determine whether the colony-lift assay significantly influenced the apparent fixative-resistant phenotype of *GUS*<sup>AR</sup>, activity assays were also carried out in cell extracts. The *GUS*-deficient strain, pREP4/GMS407, was transformed with vectors that expressed either wild-type *gusA*, the *GUS*<sup>AR</sup> variant, or the *lacZ*  $\alpha$ -fragment. Cell extracts from induced cultures were treated with glutaraldehyde or formaldehyde for 20 min at 23°C, and diluted 100-fold in buffer containing saturating concentrations of a GUS substrate, p-nitrophenyl  $\beta$ -D-glucuronide (PNPG). The extracts containing wild type or *GUS*<sup>AR</sup> catalyzed the hydrolysis of PNPG; no hydrolysis was detected in the negative control extracts, in which only the *lacZ*  $\alpha$ -fragment was expressed (data not shown). Treatment of the extract containing wild-type GUS with only 0.04% glutaraldehyde for 20 min at 23°C reduced catalytic activity by  $99.6 \pm 0.24\%$ . In sharp contrast, the *GUS*<sup>AR</sup> extract retained  $78.1 \pm 0.69\%$  of its activity after treatment with a fivefold higher (0.2%) concentration of glutaraldehyde (Fig. 3A).

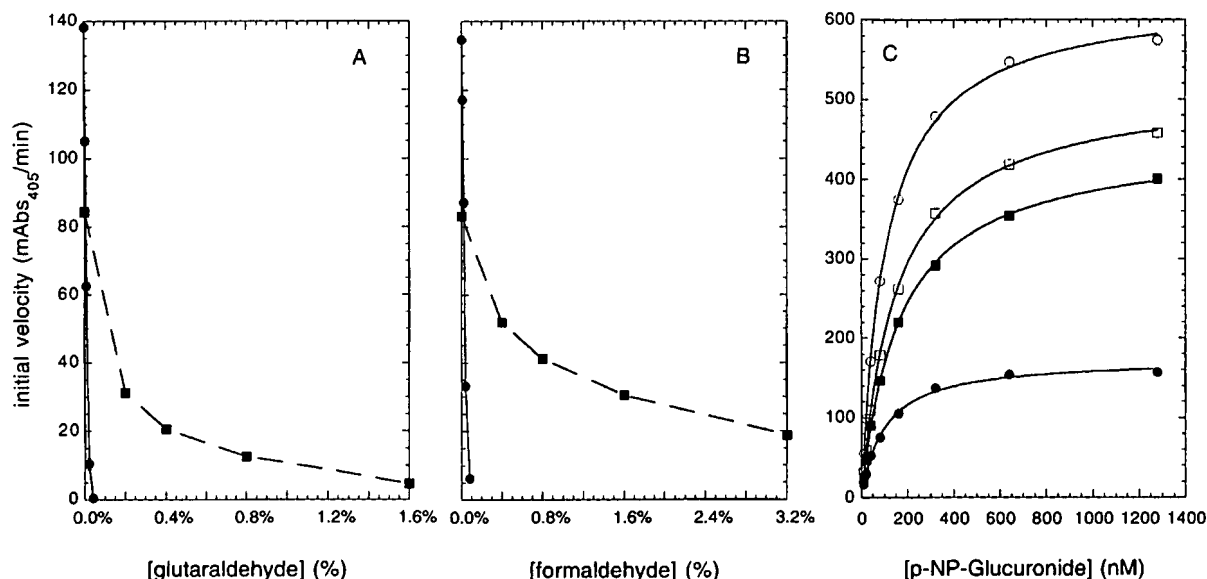
## [Glutaraldehyde]



**Figure 2.** Detection of glutaraldehyde-resistant GUS activity. *Escherichia coli* cells transformed with vectors expressing the wild-type (A, B left), a pool of the ten glutaraldehyde-resistant variants from round 1 (B right, C left) or the most resistant variants from rounds 2 (C right, D left), or 3 (D right) were streaked onto noninducing plates. The colonies were propagated, induced, and treated for 20 min with the indicated concentrations of glutaraldehyde, then reacted with X-gluc and NBT, as described in the legend to Figure 1.

Extracts were also separately treated with formaldehyde to assess whether the fixative-resistant phenotype was specific to glutaraldehyde. Again, the *GUS*<sup>AR</sup> variant exhibited much greater resistance to the fixative than did the wild-type GUS. The wild-type extract retained only  $4.6 \pm 0.05\%$  of its catalytic activity after treatment with 0.08% formaldehyde; the *GUS*<sup>AR</sup> extract retained  $62.4 \pm 0.40\%$  activity after incubation with 0.4% formaldehyde (Fig. 3B). To determine how sequence and chemical modifications may have influenced GUS activity, we conducted kinetic studies of the mutant enzyme (Fig. 3C). The wild-type and evolved *gus* genes were subcloned, expressed as fusion proteins with N-terminal hexahistidine tags, and purified by immobilized metal ion adsorption chromatography. Purified enzymes were assayed with varying concentrations of PNPG (Fig. 3C). The kinetic parameters of the wild-type enzyme ( $K_M$  for the complex with PNPG =  $110 \pm 2.9 \mu\text{M}$ ;  $k_{cat}$  =  $920 \pm 7.3 \text{ s}^{-1}$ ) were very similar to those of the *GUS*<sup>AR</sup> variant ( $K_M$  =  $150 \pm 3.9 \mu\text{M}$ ;  $k_{cat}$  =  $750 \pm 5.8 \text{ s}^{-1}$ ). The kinetic parameters of wild-type and evolved enzymes were also determined following reaction with a sublethal concentration (0.04%) of formaldehyde for 20 min at 23°C. Formaldehyde had a larger effect on the turnover number ( $250 \pm 17 \text{ s}^{-1}$  for the partially modified wild type,  $650 \pm 4.7 \text{ s}^{-1}$  for the modified *GUS*<sup>AR</sup> variant), than on the Michaelis constants ( $99 \pm 6.7 \mu\text{M}$  and  $170 \pm 3.7 \mu\text{M}$  for the modified wild-type and *GUS*<sup>AR</sup> enzymes,

## RESEARCH



**Figure 3.** GUS catalytic activity as a function of aldehyde concentration. *Escherichia coli* cell extracts containing the wild-type (circles) or evolved GUS<sup>AR</sup> (squares)  $\beta$ -glucuronidase were incubated for 20 min at 23°C in buffer containing the indicated concentrations of glutaraldehyde (A) or formaldehyde (B). The protein was then diluted 100-fold into buffer containing the chromogenic GUS substrate, p-nitrophenyl- $\beta$ -D-glucuronide (PNPG). Hydrolysis of the substrate was followed at 405 nm using a spectrophotometer (see Experimental Protocol). Each point represents the average of three initial velocity values; the points subsume the error bars. The control extract from an isogenic strain not expressing GUS does not have detectable activity (not shown). (C) Purified wild-type (circles) or evolved GUS<sup>AR</sup> (squares) enzymes were incubated for 20 min at 23°C in buffer containing 0% (empty symbols) or 0.04% (filled symbols) formaldehyde, then reacted with the indicated concentrations of PNPG. The initial velocity values were fitted to the Michaelis-Menten equation (lines); the derived kinetic parameters are presented in the text.

respectively). These results in conjunction with the cell extract data show that GUS<sup>AR</sup> is expressed at lower levels than the wild type, but is inherently more aldehyde resistant.

Sequence and structural mapping of the evolved GUS<sup>AR</sup> variant. Upon sequencing, the evolved *gusA* gene was found to contain the following amino acid substitutions: N66D, D151N, A219V, I396T, T480A, Q498R, D508E and K567R, as well as six silent mutations. Only one of these changes, the D508E mutation, results from a transversion, indicating a strong transition bias in our random mutagenesis methods. The amino acid sequences of the *E. coli* and human GUS proteins are 50% identical<sup>18</sup> and could be readily aligned by the algorithm devised by Needleman and Wunsch<sup>19</sup> using the program GAP 4.0 (Genetics Computing Group, Madison, WI). Both proteins are tetramers<sup>16,18</sup> and are virtually identical in substrate specificity<sup>1</sup>. The positions of the loci that were altered in the evolved *E. coli* enzyme could thus be tentatively mapped onto the crystal structure of the human GUS protein<sup>18</sup> (Fig. 4).

The 10 *gusA* mutants isolated in the first round of screening were sequenced; mutations were found at a frequency of three per 1.8 kb. Seven of the first-round variants contained amino acid substitutions (K567R, T480A, D508E or N66D/D151N) that were subsequently found in the most active third-round GUS<sup>AR</sup> variant. It is instructive that many of the single substitutions that confer modest resistance to aldehyde modification can interact additively or synergistically to confer robust resistance to aldehyde modification.

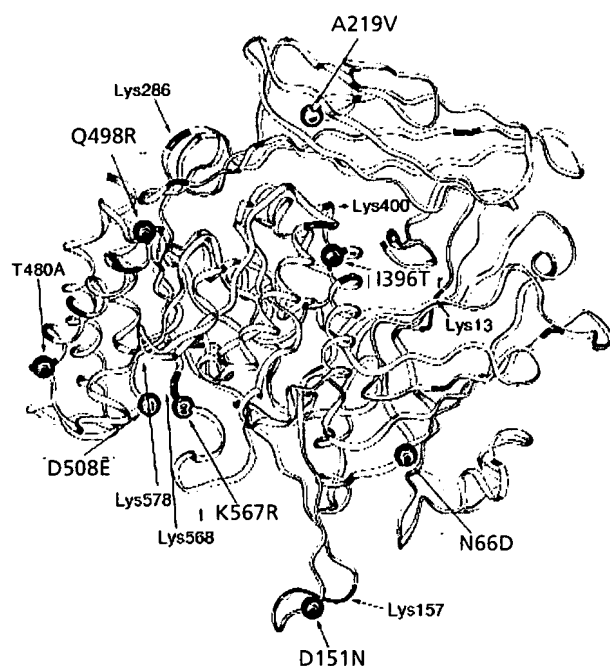
GUS<sup>AR</sup> as a lineage tracer in *Xenopus* embryos. The N358S mutant of GUS is a commercially available and commonly used reporter gene in plants<sup>4</sup>. The N358S mutation eliminates a cryptic glycosylation site, and should not affect its function in the cytoplasm; we chose this construct because it also contained the upstream sequences necessary for expression in eukaryotic cells<sup>20</sup>. In order to determine if the fixative-resistant GUS might also prove useful in other model organisms, transcripts encoding

GUS<sup>AR</sup> and N358S GUS were microinjected into 16- to 32-cell stage *Xenopus* embryos. Two days later, the embryos were fixed in 3.7% formaldehyde for 20 min and stained using a standard protocol for the detection of *lacZ* expression, except that X-gluc was substituted for X-gal. The descendants of cells injected with the wild-type-like N358S GUS mRNA did not change color (Fig. 5A). Endogenous GUS activity was apparently also abrogated by the 20 min incubation in 3.7% formaldehyde. In contrast, the descendants of cells injected with the GUS<sup>AR</sup> mRNA turned bright blue-green (Fig. 5B).

In order to determine if multiple reporters might be used in tandem for lineage analysis, mRNAs encoding either N358S or the GUS<sup>AR</sup> were co-injected into embryonic cells along with mRNA encoding *lacZ*. When the embryos were first stained with X-gluc, again only the cells that inherited the GUS<sup>AR</sup> mRNA turned blue-green (Fig. 6A and B). The embryos were subsequently stained with rose-gal, a  $\beta$ -galactosidase substrate that forms a red precipitate. In embryos that received either the N358S or GUS<sup>AR</sup>, some cells were colored red, indicating the inheritance of *lacZ*. However, in embryos that received GUS<sup>AR</sup> some cells or patches were also purple, indicating the co-inheritance of the GUS<sup>AR</sup> and *lacZ* (Fig. 6C and D). The *lacZ* mRNA in this experiment also served as an internal control that demonstrated that RNAs were entering cells and surviving until fixation.

## Discussion

**Mechanisms of aldehyde resistance.** The *E. coli* GUS protein contains 27 lysine residues, six of which are conserved among the sequenced GUS genes. Although the fraction of lysine residues that are modified by aldehydes is unknown, wild-type GUS activity is quite susceptible to even low levels of fixatives. For example, the catalytic activities of the wild-type (Fig. 3B) and N358S GUS (Figs 5A and 6A) are inactivated by <3.7% formaldehyde, while  $\beta$ -galactosidase, which contains 20 lysine residues, is not (Fig. 6). Taken together, these results suggest that one or more of the GUS



**Figure 4.** Homology mapping of amino acid substitutions that confer aldehyde resistance. The crystal structure of a subunit of the C $\alpha$  trace of human GUS<sup>18</sup> is shown. The amino acid sequences of the *E. coli* and human GUS proteins were aligned using the application GAP 4.0 (Genetics Computer Group) and were found to be 48.5% identical. The positions of lysine residues are darkened, and the conserved lysines are labeled. The positions of amino acid substitutions in the evolved GUS<sup>AR</sup> *E. coli* protein are shown as balls.

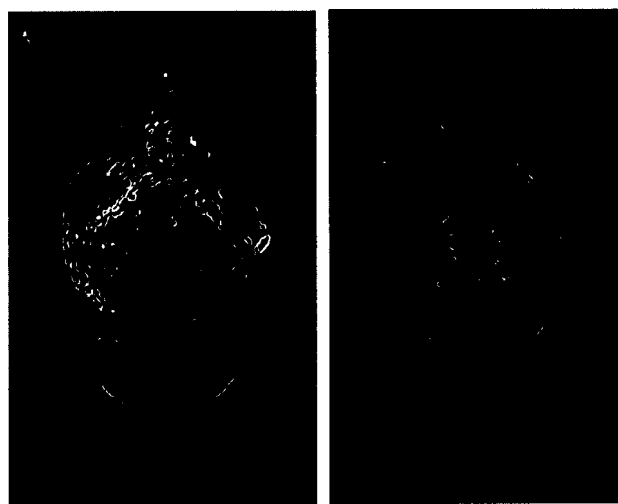
lysine residues is either itself critical for activity or presents a conjugation site that leads to functional disruption. However, identifying which of the many lysine residues in GUS were responsible for inhibition by fixatives would have been a daunting task. Instead, we relied on a random mutagenesis to identify GUS variants with catalytic activity resistant to aldehydes. Following three rounds of screening and amplification, we isolated an octuple-mutant GUS<sup>AR</sup> with catalytic activity resistant to roughly 80-fold higher levels of glutaraldehyde than the wild-type activity (Fig. 3A).

Surprisingly, only one of the amino acid substitutions, K567R, in the evolved GUS<sup>AR</sup> occurred at a lysine residue. Since AAA or AAG encodes lysine, the apparent transition bias in our random-mutagenesis method and the size of our initial library provided ample opportunities for each lysine to conservatively mutate into arginine (AGA or AGG). While it is possible that mutation of this single lysine was largely responsible for protection against aldehydes, this explanation is unlikely. Three of the 10 clones isolated in the first round of screening contain the K567R sequence substitution, but none of these first-round isolates are as resistant as any of the second-round isolates (Fig. 2C). The aldehyde resistance of GUS progressively increased over three rounds of screening and selection, and the final product had accumulated seven additional amino acid substitutions. The finding that amino acid substitutions that modulate protein function are dispersed in the primary and tertiary structure of GUS is congruent with previous attempts to evolutionarily engineer the physical and kinetic parameters of enzymes. Experiments that directed an increase in the catalytic activity of a p-nitrobenzyl esterase in organic enzymes yielded multiple sequence substitutions scattered throughout the tertiary structure<sup>21</sup>. Site-directed mutation studies of T4 lysozyme have shown that stabilizing amino acid changes, which occur in the core of that enzyme, are additive in effect<sup>22</sup>.

Interestingly, the seven non-lysine amino acid substitutions mapped onto the surface of the protein near lysine residues (Fig. 4). Protein structure is more highly conserved than protein sequence<sup>23</sup>, and since the primary sequences of the *E. coli* and human GUS enzymes are quite similar (48.5% identity<sup>8</sup>), it can be conservatively assumed that their tertiary structures also align well. Based on this assumption, we can advance hypotheses regarding the contributions of individual amino acid substitutions to aldehyde resistance. For example, Lys568 (*E. coli* numbering) is conserved among the sequenced GUS genes, and is in the active-site<sup>18</sup>. The C $\alpha$ -C $\alpha$  distance from Lys568 to the D508E substitution is 3.97 Å (Fig. 4). Since the lysine side chain is 7 Å in length, this adjacent sequence substitution might raise the pK<sub>a</sub> of the epsilon amino group of Lys568, thereby reducing its reactivity with aldehydes. Similarly, the K567R substitutions already mentioned is within 3.81 Å of the active-site lysine, and mutation to arginine may prevent modification that could sterically interfere with substrate binding or catalysis.

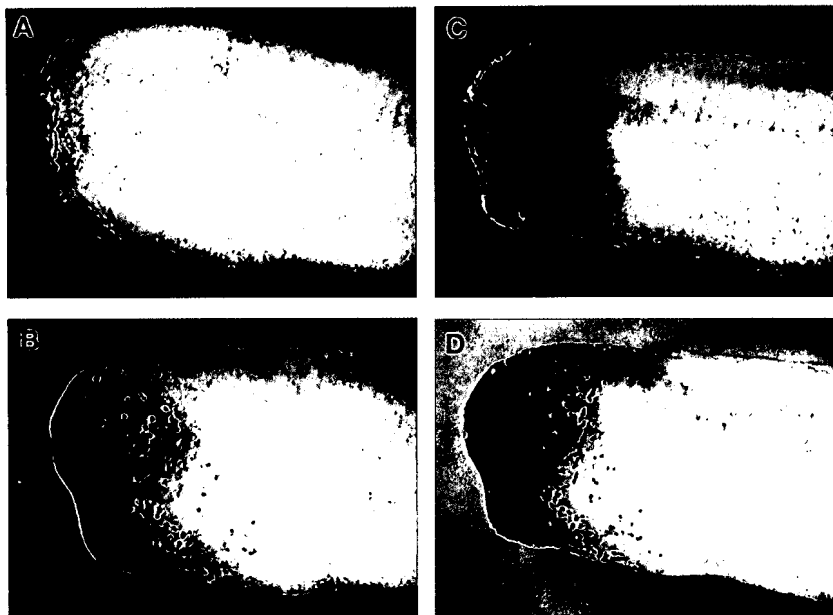
Similarly,  $\beta$ -glucuronidase is active only as a tetramer<sup>16,18</sup>, and lysines play a key role in its quaternary structure. To the extent that modifications of interfacial lysines disrupt quaternary structure and enzymatic function, adjacent amino acid substitutions could render these lysines less reactive. In this regard, the loop containing D151N and three lysines is <5 Å away from the  $\alpha$ -helix of the adjacent subunit containing T480A and three other lysines. These amino acid substitutions could also prevent structural and functional disruption by independently increasing the affinity between the subunits. For example, the A219V substitution also maps to the other interface, although it is not immediately adjacent to any lysines.

Overall, it appears the surface chemistry of the enzyme has coordinately evolved either to cause lysines to be less reactive or to functionally accommodate covalent modification of lysines. Our results suggest that there may be multiple possible routes by which proteins could be adapted to function in a wide variety of fixatives or solvent systems. More importantly, they suggest a way of augmenting protein chemistry by introducing amino acids with novel surface conjugates.



**Figure 5.** Expression of GUS<sup>AR</sup> in *Xenopus* embryos. Embryos at the 16- to 32-cell stage were injected with 1 ng of mRNA encoding N358S (A) or GUS<sup>AR</sup> (B) and fixed two days later in 3.7% formaldehyde for 20 min. GUS activity was detected using the chromogenic substrate X-gluc (light blue/green). The reddish-purple color of the cement gland of the embryo shown in (A) is from a natural pigment, and the blue color of the embryo shown in (B) is from GUS<sup>AR</sup> activity.

## RESEARCH



**Figure 6. Multiple marker staining of *Xenopus* embryos.** Embryos were co-injected with 0.5 ng of *lacZ* mRNA and 1 ng of either N358S GUS (A, C) or GUS<sup>AR</sup> (B, D), and subsequently fixed in 3.7% formaldehyde for 20 min. Following fixation, embryos were reacted with X-gluc (light blue/green in all frames). As in Figure 5, embryos injected with mRNA encoding GUS<sup>AR</sup> (B) stained much more intensely than those injected with N358S GUS mRNA (A). All embryos were subsequently rinsed free of X-gluc and stained with the chromogenic substrate for  $\beta$ -galactosidase, rose-gal (red in C and D).

GUS<sup>AR</sup> as a universal reporter gene. Since most naturally occurring  $\beta$ -glucuronidases are likely to be fixative-labile, the fixative-resistant GUS<sup>AR</sup> we have isolated should prove useful for expanding the range and power of GUS staining techniques. In addition, it should be possible to develop methods for following multiple genes or cell lineages in parallel. Such methods generally rely on protocols in which fixed tissues are reacted with antibodies conjugated to dyes or reporter enzymes (for example, see ref. 24).

Reporter genes are very commonly used in *Xenopus* as cell lineage tracers, and have proved important for gene expression studies in developing embryos<sup>25,26</sup>. Following mRNA microinjection, the fixative-resistant GUS could be specifically followed in *Xenopus* relative to both background activity and the wild-type reporter. Moreover, a lineage trace in tandem with  $\beta$ -galactosidase demonstrated the use of GUS in a multiple-reporter format. These experiments pave the way for the practical development of two-enzyme reporter systems, and could potentially be combined with a  $\beta$ -lactamase reporter system developed by Raz *et al.*<sup>27</sup> to create three-enzyme reporter systems.

The results in *Xenopus* embryos are notable in that that no special precautions were taken to enhance gene expression or enzymatic activity. In contrast to the transformation of reporter constructs, microinjected reporter mRNAs do not replicate and their dosage progressively decreases as messages are segregated or broken down. Further, no attempt was made to increase the signal intensity of GUS<sup>AR</sup> by fusing it to a nuclear localization signal, as was the case for the *lacZ* control. Nor were fluorescent or other highly sensitive commercially available GUS substrates<sup>3</sup> utilized. In short, the evolved enzyme is itself robust enough so that new staining techniques can easily be adapted from extant methods.

### Experimental protocol

**Materials.** DNA-modifying enzymes, including restriction enzymes and Vent polymerase, were purchased from New England Biolabs (Beverly, MA). Deoxyribonuclease I was from GIBCO-BRL (Gaithersburg, MD).

Taq polymerase was expressed and purified as described by Grimm and Arbutnot<sup>28</sup>. DNA sequencing kits were from Perkin-Elmer/Applied Biosystems (Foster City, CA). Cloning vector pGEM-5 was from Promega (Madison, WI), pBluescript II SK(+) was from Stratagene (La Jolla, CA), and the regulatory vector pREP4 was from Qiagen (Chatsworth, CA). pGUS N358S was from Clontech (Palo Alto, CA), and pET28a(+) from Novagen (Madison, WI). DNA purification columns were purchased from Qiagen (Chatsworth, CA). X-gluc was from Gold Biotechnology (St. Louis, MO) and Butterfly nitrocellulose membranes from Schleicher and Schuell (Keene, NH). The mMessage mMachine SP6 in vitro mRNA transcription kit was from Ambion (Austin, TX). MicroSpin G-25 Sephadex spin columns were from Pharmacia Biotech (Piscataway, NJ). *Escherichia coli* strain InvαF<sup>+</sup> was from Invitrogen (Carlsbad, CA), W3110 (ATCC No. 27325) from the ATCC (Rockville, MD), GMS407 from the *E. coli* Genetic Stock Center (New Haven, CT), and BL21(DE3)pLysS from Novagen. Other chemicals, including glutaraldehyde, PNPG and NBT, were from Sigma Chemicals (St. Louis, MO).

**Cloning of *gusA*.** The *E. coli gusA* gene was amplified from W3110 cells using Vent polymerase and the primers 5'-CCGGATCCTCTAGACATCT-TACGTCCTGTAGAAACC-3' and 5'-GCCAATTCGTCAGTCATTGTTTGCCTCCCTGCT-3' (*Xba*I and *Eco*RI sites underlined). The PCR product was blunt-end ligated into the *Eco*RV site of pGEM-5 by standard methods<sup>29</sup>. *Escherichia coli* InvαF<sup>+</sup> cells were transformed by the method described by Inoue *et al.*<sup>30</sup>. The *gusA* gene was subcloned into pBluescript II SK(+) using restriction endonucleases *Xba*I and *Eco*RI. The nucleotides encoding the *lacZ* α-fragment that would normally have been located between the ribosome binding site and the *gusA* start codon were deleted by amplifying the remainder of the plasmid using primers 5'-CCGGATCCTC-TAGACATCTTACGTCCTGTAGAAACC-3' and 5'-CGTCTAGAAGCT-GTTTCTCTGTGTAATTG-3', digesting with *Xba*I, ligating, and transforming pREP4/InvαF<sup>+</sup> (see below). The resultant construct placed the *gusA* gene under direct control of the *lac* promoter. The GUS expression vector was named *gusA*-pBSA.

**Library construction and screening.** For the first round of screening, random mutations were introduced into the cDNA by mutagenic PCR<sup>15</sup> using primers 5'-CCAGTCACGACGTTGTAAAA CGACG-3' and 5'-ATGCTTC-CGGCTCGTATGTTGTGTGG-3', which anneal to the pBluescript II SK(+) vector outside of the boundaries of the *gusA* insert. The amplification reaction was carried out with 100 nM primers, 60 mM Tris-HCl pH 8.5, 15 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.2 mM MgCl<sub>2</sub>, 0.125 mM MnCl<sub>2</sub>, 0.2 mM dGTP, 0.2 mM dATP, 0.4 mM dTTP, 0.4 mM dCTP, for 35 cycles of 94°C × 30 s, 72°C × 2 min. The *gusA*-pBSA plasmid library was transformed into *E. coli* InvαF<sup>+</sup> cells harboring the *lacI* expression vector pREP4. The plasmid was unstable when propagated in *E. coli* InvαF<sup>+</sup> without pREP4, probably because the *lac* repressor is not present at high enough levels to limit expression of the *gusA*. For colony-lift assays, *gusA*-pBSA/pREP4/InvαF<sup>+</sup> colonies were propagated on liquid Luria Broth supplemented with 25 μg/ml kanamycin and 100 μg/mL ampicillin (LB-kan/amp) + 0.4% glucose plates for 12 h at 37°C. The colonies adsorbed to a nitrocellulose filter and were transferred colony side up to LB-kan/amp plates containing 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG), and induced at 37°C for 12–24 h. The nitrocellulose-bound colonies were transferred to GUS buffer (50 mM sodium phosphate pH 7.0, 0.1% Triton X-100, 1 mM EDTA) containing 0.2% glutaraldehyde and incubated for 20 min at 23°C. The filters were then transferred to buffer containing 165 μg/ml X-gluc and 330 μg/ml NBT and incubated for 10–30 min. The filters were incubated briefly in 4 M guanidine hydrochloride to arrest color development. Those colonies on the master plate corresponding to the darkest colony remnants on the filter were isolated and amplified.

For subsequent rounds of screening, the alleles were randomly recombined and mutated by DNA shuffling as described by Stemmer<sup>17</sup>. In short, the *gusA* variants were PCR amplified using the same primers as in the

mutagenic PCR reactions already described, partially digested with DNase I, and reassembled in a PCR reaction without primers. The products were amplified in a PCR with primers, then subcloned back into *gusA*-pBSA for screening. The second and third rounds were carried out in the same way, except that 1.0% and 3.5% glutaraldehyde was used to fix the colonies before the incubation in X-gluc and NBT. The most resistant round 3 clone was isolated and sequenced at the University of Texas, Institute of Cellular and Molecular Biology Core Facility using the Applied Biosystems protocol, via the primers originally used for mutagenic PCR and two additional internal primers: 5'-CGCCGGGAATGCTGATTACC-3' and 5'-CTGATGGTATCGCT-GTGACGC-3'.

**In vitro characterization of enzyme activity.** For the preparation of lysates, *gusA*-pBSA/pREP4/GMS407 cells were propagated at 37°C in LB-kan/amp. The *gusA* gene was induced by the addition of 0.5 mM IPTG to mid-log ( $OD_{600} = 0.3$ ) cultures, and the induced cultures were grown overnight. Cells were centrifuged, resuspended in distilled water, centrifuged again, and resuspended in GUS buffer. Cells were lysed with the addition of 10 mM EDTA and 1 mg/ml chicken lysozyme. The insoluble fraction was centrifuged down, and the aldehyde resistance of the GUS in the supernatant was determined as follows. Glutaraldehyde or formaldehyde was added to an aliquot of supernatant and the mixture was incubated at 23°C for exactly 20 min. The mixture was then diluted 1/100 into GUS buffer containing 0.5 mM PNPG. The hydrolysis of the substrate was followed for 1 min at 23°C at 405 nm in a Shimadzu UV-1601 spectrophotometer. The absorption extinction coefficient of p-nitrophenol under these conditions was  $11.50 \text{ mM}^{-1} \text{ cm}^{-1}$ . The initial rates of hydrolysis were linear (data not shown).

To generate purified GUS enzymes, the wild-type and evolved *gusA* genes were amplified by PCR with the primers: 5'-GCTCTAGACGATATGT-TACGTCCTGTAGAAACC-3' and 5'-GCCAATCTGTCAGTCATTGTTTGC-CTCCCTGCT-3' and subcloned into the expression vector pET28a(+) using the restriction enzymes *NdeI* and *EcoRI* (sites underlined in primers). The resultant genes were sequenced as described already to confirm that no additional mutations had been introduced during amplification or cloning. The expression constructs were transformed into BL21(DE3)/pLysS. The transformed strains were propagated and induced, and the proteins purified by nickel chelate chromatography, as suggested by Novagen (Madison, WI). The protein preparations were judged to be >99% pure following SDS-PAGE and Coomassie Blue staining (data not shown). Purified protein concentrations were determined via Bradford protein assays (Bio-Rad, Hercules, CA).

A 10 pmol quantity of purified GUS protein was preincubated for 20 min at 23°C in 10  $\mu\text{l}$  of GUS buffer (1  $\mu\text{M}$ ) containing 0% or 0.04% formaldehyde. Then, 5  $\mu\text{l}$  of protein solution were added to 1 ml of buffer (5 nM) containing varying concentrations of PNPG, and the initial velocity of each reaction was determined as already described. The kinetic parameters of the wild-type and mutant enzymes were calculated by fitting the initial velocity values to the Michaelis-Menten equation using the application Kaleidagraph 3.0.5 (Adelbeck Software, Reading, PA).

**Expression of *gusA* in *Xenopus* embryos.** The *GUS*<sup>AR</sup> gene was subcloned into pGUS N358S; this placed the gene downstream of a Kozak sequence<sup>20</sup>, so that its transcript could be recognized by eukaryotic translation systems. N358S *GUS* and *GUS*<sup>AR</sup> were subcloned into the *Xenopus* expression vector p64TS. This plasmid provides in vitro-transcribed mRNAs with *Xenopus* globin 5'- and 3'-untranslated regions and greatly increases the amount of protein translated from the mRNA<sup>31</sup>. Capped mRNA was produced by *in vitro* transcription<sup>32</sup> of the clones already described using the Ambion mMessage mMaker SP6 protocol. In vitro transcriptions were also treated with DNase I, and the mRNA was purified using a Sephadex G-25 spin column to minimize nonspecific toxicity effects. Purified mRNAs were resuspended in sterile water for injections.

Female adult *Xenopus* were induced to ovulate with human chorionic gonadotropin, and eggs were fertilized in vitro. Embryos were dejellied in 3% cysteine solution and washed in 0.2 $\times$  MMR<sup>33</sup>. Embryos were then reared at 13–18°C in 0.2 $\times$  MMR. Microinjections were performed as described<sup>34</sup>. Embryos were fixed in MEMFA (0.1 M MOPS, pH 7.4 / 2 mM EGTA / 1 mM  $\text{MgSO}_4$  / 3.7% formaldehyde) for 20 min, and embryos were washed 5  $\times$  5 min in 1 $\times$  PBS. GUS activity was detected using 1 mg/ml X-gluc in a solution of 1 $\times$  PBS / 20 mM potassium ferriicyanide / 20 mM potassium ferrocyanide / 2 mM  $\text{MgCl}_2$  / 0.02% NP-40 at 37°C for 2 h.  $\beta$ -Galactosidase was detected using the same buffer but substituting rose-gal for X-gluc. Injection experiments repeated on different days with different preparations of mRNA gave similar results (data not shown).

## Acknowledgments

We thank the Office of Naval Research for funding. I.M. was supported by a National Science Foundation/Alfred P. Sloan Postdoctoral Research Fellowship in Molecular Evolution (DBI-9750002). We thank Dr. Mary Berlyn of the E. coli Genetic Stock Center for sending us strain GMS407 and Ms. Sabine Bell for synthesizing oligonucleotides. Finally, we thank members of the Ellington group and Dr. William Wu of Ambion for helpful discussion.

- Jefferson, R.A., Burgess, S.M. & Hirsh, D. Beta-glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* **83**, 8447–8451 (1986).
- Martin, T., Woehner, R.V., Hummel, S., Willmitzer, L. & Frommer, W.B. in *GUS Protocols: using the GUS gene as a reporter of gene expression*. (ed. Gallagher, S.R.), 23–43 Academic Press, New York; 1992).
- Naleway, J.J. in *GUS Protocols: using the GUS gene as a reporter of gene expression*. (ed. Gallagher, S.R.), 61–76 (Academic Press, New York; 1992).
- Farrell, L.B. & Beachy, R.N. Manipulation of beta-glucuronidase for use as a reporter in vacuolar targeting studies. *Plant Mol. Biol.* **15**, 821–825 (1990).
- Craig, S., in *GUS Protocols: using the GUS gene as a reporter of gene expression* (ed. Gallagher, S.R.), 115–124 (Academic Press, New York; 1992).
- Jefferson, R.A. The GUS reporter gene system. *Nature* **342**, 837–838 (1989).
- Habeeb, A.F.S.A. & Hiramoto, R. Reactions of proteins with glutaraldehyde. *Arch. Biochem. Biophys.* **126**:16–26 (1968).
- Schlarman, H.R., Risseeuw, E., Franke-van Dijk, M.E. & Hooykaas, P.J. Nucleotide sequence corrections of the *uidA* open reading frame encoding beta-glucuronidase. *Gene* **138**, 259–260 (1994).
- Oshima, A. et al. Cloning, sequencing, and expression of cDNA for human beta-glucuronidase. *Proc. Natl. Acad. Sci. USA* **84**, 685–689 (1987).
- D'Amore, M.A., Gallagher, P.M., Korfhagen, T.R. & Ganschow, R.E. The complete sequence and organization of the murine beta-glucuronidase gene. *Biochemistry* **27**, 7131–7140 (1988).
- Nishimura, Y. et al. Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and *in vitro* insertion of its encoded polypeptide into microsomal membranes. *Proc. Natl. Acad. Sci. USA* **83**, 7292–7296 (1986).
- Ray, J. et al. Cloning of canine beta-glucuronidase cDNA, mutation identification in canine MPS VII, and retroviral vector-mediated correction of MPS VII cells. *Genomics* **48**, 248–253 (1988).
- Zhang, J.H., Dawes, G. & Stemmer, W.P. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. USA* **94**, 4504–4509 (1997).
- Giver, L., Gershenson, A., Freskgard, P.O. & Arnold, F.H. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. USA* **95**, 12809–12813 (1998).
- Cadwell, R.C. & Joyce, G.F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
- Blanco, C. & Nemoz, G. One step purification of *Escherichia coli* beta-glucuronidase. *Biochimie* **69**, 157–161 (1987).
- Stemmer, W.P.C. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751 (1994).
- Jain, S., Drendel, W.B., Chen, Z.W., Mathews, F.S., Sly, W.S. & Grubb, J.H. Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat. Struct. Biol.* **3**, 375–381 (1996).
- Needleman, S.B. & Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
- Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148 (1987).
- Moore, J.C. & Arnold, F.H. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotechnol.* **14**, 458–467 (1996).
- Zhang, X.J., Baase, W.A., Shoichet, B.K., Wilson, K.P. & Mathews, B.W. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng.* **8**, 1017–1022 (1995).
- Holm, L. & Sander, C. Mapping the protein universe. *Science* **273**, 595–603 (1996).
- van der Loos, C.M., Becker, A.E. & van den Oord, J.J. Practical suggestions for successful immunoenzyme double-staining experiments. *Histochem. J.* **25**, 1–13 (1993).
- Wallingford, J., Seufert, D., Virta, V. & Vize, P. p53 Activity is essential for normal development in *Xenopus laevis*. *Curr. Biol.* **7**, 747–757 (1997).
- Wallingford, J., Carroll, T. & Vize, P. Precocious expression of the Wilms' tumor gene *xWT1* inhibits embryonic kidney development in *Xenopus laevis*. *Dev. Biol.* **202**, 103–112 (1998).
- Raz, E., Zlokarnik, G., Tsien, R.Y. & Driever, W. Beta-lactamase as a marker for gene expression in live zebrafish embryos. *Dev. Biol.* **203**, 290–294 (1998).
- Grimm, E. & Arbutnot, P. Rapid purification of recombinant Taq DNA polymerase by freezing and high temperature thawing of bacterial expression cultures. *Nucleic Acids Res.* **23**, 4518–4519 (1995).
- King, P.V. & Blakesley, R.W. Optimizing DNA ligations for transformations. *FOCUS* **8**, 30–32 (1986).
- Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
- Krieg, P.A. & Melton, D.A. Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned cDNAs. *Nucleic Acids Res.* **12**, 7057–7070 (1984).
- Krieg, P. & Johnson, A. in *A laboratory guide to RNA: isolation, analysis, and synthesis*. (ed. Krieg, P.) 141–153 (Wiley-Liss, New York; 1996).
- Peng, H.B. Solutions and protocols. *Methods Cell Biology* **36**, 657–662 (1991).
- Vize, P., Melton, D., Hemmati-Brivanlou, A. & Harland, R. Assays for gene function in developing *Xenopus* embryos. **36**, 367–387 (1991).



## INVITED EDITORIAL

# Genomic Sequence, Splicing, and Gene Annotation

Stephen M. Mount

Department of Cell Biology and Molecular Genetics, University of Maryland, College Park

### Introduction

The sequence of the human genome is at hand. Most scientists who use the sequence will rely on annotations that provide information about the number and location of genes and about their inferred protein products. Traditionally, genes have been annotated by scientists with a particular interest in them. However, annotation of the complete human genome sequence will have to be at least partially automated. Gene annotation incorporates cDNA data (including expressed sequence tags [ESTs]), sequence similarity, and computational predictions based on the recognition of probable splice sites and coding regions (Stormo 2000; also see David Haussler's Web site, Computational Genefinding). The state of the art was recently surveyed by the Genome Annotation Assessment Project-GASP1 and must be regarded as imperfect (Bork 2000; Reese et al. 2000).

This review enumerates aspects of pre-mRNA splicing that limit our ability to predict gene structure from genomic sequence, drawing on the recently annotated complete genome of *Drosophila melanogaster* (Adams et al. 2000) as an example. In particular, the following four facts will be discussed. First, splice sites do not always conform to consensus. Second, noncoding exons are common. Third, internal exons can be arbitrarily small, and small internal exons confound not only gene finding but also the alignment of cDNA and genomic sequences. Fourth, splice sites are not recognized in isolation, and nucleotides that are far from splice sites can affect splicing. This list and the accompanying analysis should make molecular geneticists aware of the ways in which gene annotations can be wrong and should encourage recourse to the primary data. In addition, the same considerations indicate that inherited disease can

be caused by mutations remote from splice sites that nevertheless affect splicing.

### Discussion

#### *Splice Sites Do Not Always Conform to Consensus*

It is well established that nearly all splice sites conform to consensus sequences (Mount 1982; Senapathy et al. 1990; Zhang 1998). These consensus sequences include nearly invariant dinucleotides at each end of the intron—GT at the 5' end of the intron and AG at the 3' end of the intron. Most gene-finding software and most human annotators will find only introns that begin with a GT and end with an AG. However, nonconsensus splice sites have been described, and I will discuss three classes, in decreasing order of frequency.

The most common class of nonconsensus splice sites consists of 5' splice sites with a GC dinucleotide. Senapathy et al. (1990) listed 17 examples among 3,724 5' splice sites, suggesting a frequency of ~0.5%. Jackson (1991) listed a total of 26 GC sites, whereas Wu and Krainer (1999) cited an additional 18 examples. GC 5' splice sites are consistent with the experimental observation that, of the six possible point mutations within the GT dinucleotide, mutation of T to C in position 2 has the smallest effect on in vitro splicing (Aebi et al. 1986). At other positions within the consensus, GC sites conform extremely well to the standard consensus; for example, 42 of the 44 sites cited above have a consensus G residue at both position -1 and position +5. It is reasonable to assume that GC sites are recognized by the standard (U2-dependent) spliceosome.

The second class of exception to splice-site consensus is U12 introns, a minor class of rare introns with splice-site sequences that are very different from the standard consensus but that are very similar to each other. The existence of this class was first pointed out by Jackson (1991) and was considered in more detail by Hall and Padgett (1994). It was subsequently discovered that U12 introns are removed by a minor spliceosome containing the rare U11, U12, U4atac, and U6atac snRNPs, in place of U1, U2, U4, and U6 (Tarn and Steitz 1997; Burge et al. 1998). Some U12 introns have AT and AC in place of GT and AG and are known as "AT-AC" introns. However, terminal intron dinucleotide sequences do not

Received August 3, 2000; accepted for publication August 15, 2000; electronically published September 8, 2000.

Address for correspondence and reprints: Dr. Stephen M. Mount, Department of Cell Biology and Molecular Genetics, H. J. Patterson Hall, University of Maryland, College Park, MD 20742-5815. E-mail: sm193@umail.umd.edu

This article represents the opinion of the author and has not been peer reviewed.

© 2000 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2000/6704-0003\$02.00

distinguish between U2- and U12-dependent introns (Dietrich et al. 1997). Rather, U12 introns can be identified by highly conserved sequences at the 5' splice site (RTATCCTY; R = A or G, and Y = C or T) and branch site (TCCTRAY). U12 introns are found in many eukaryotes, including *Drosophila melanogaster* (Adams et al. 2000) and *Arabidopsis thaliana* (Shukla and Padgett 1999) but not *Caenorhabditis elegans*.

Finally, there are a small number of nonconsensus sites that fit into neither of the two categories mentioned above. Many reports of such variant splice sites can be traced to errors in annotation or interpretation, polymorphic differences between the sources of cDNA and genomic sequence, inclusion of pseudogene sequences, or failure to account for somatic mutation (author's unpublished data; for examples, see Jackson 1991). However, there are many examples of sites that match the consensus very poorly, and experimental work has established that 5' splice sites do not absolutely require GT—and that 3' splice sites do not absolutely require AG—in order to be recognized in vivo (Aebi et al. 1986; Roller et al. 2000, and references therein). In yeast, an intron that is within the HAC1 mRNA and that has no similarity to the standard nuclear pre-mRNA intron consensus sequence is spliced by a specific, regulated, endonuclease and tRNA ligase (Sidrauski et al. 1996). This intron provides a precedent for introns in protein-coding genes with completely novel splice sites.

### Noncoding Exons Are Common

There is considerable confusion between exons and coding regions. The term "exon" was coined by Gilbert (1978) to refer to what is left when introns are removed by splicing, and RNAs that are entirely noncoding (such as tRNAs) are sometimes spliced. However, the term exon is often misused to refer to a stretch of coding information. In reality, however, noncoding exons are quite common, occurring in >35% of human genes (Zhang 1998). Gene-finding software generally detects sequence features characteristic of coding regions rather than of exons and does not even attempt to identify noncoding exons, or noncoding portions of exons. This is because the statistical biases introduced by protein-coding are in fact a very powerful tool for the identification of coding DNA, and no similar tool has been developed for the identification of noncoding exons.

A similar problem can arise in genes without noncoding exons. If the first intron occurs near the initiator AUG, then the coding information in the first exon can be very short and difficult to identify by measures of coding tendency. Furthermore, the first intron tends to be longer than average (Maroni 1996), and such an arrangement can separate promoter function (perhaps including downstream transcriptional enhancer elements

lying in the first intron) from the bulk of the coding information downstream. In these cases, investigators have no way of knowing how much information is missing—or where the 5' end of the gene is likely to reside—without experimental data such as a cDNA sequence or a 5' EST.

### Internal Exons Can Be Arbitrarily Small

A less frequent but perhaps more serious problem for gene-discovery methods is posed by small internal exons. Vertebrate internal exons have an average size of ~130 nucleotides (Hawkins 1988; Zhang 1998), and roughly 65% of internal human exons are 68–208 nucleotides in length (Maroni 1996). This size distribution reflects a functional constraint. Optimal splicing efficiency requires exons with sizes of ~50–300 nucleotides (Roberson et al. 1990; Dominski and Koe 1991; see review by Berget 1995). However, a considerable number, >10%, of exons are <60 nucleotides in length, and it is these exons that can be difficult to identify by measures of coding tendency.

Just how small can internal exons be? There appears to be no lower limit, and many cases of exons <10 nucleotides have been described (for examples, see Stamm et al. 1994; also see the author's Web site, Gene Annotation and Splice Site Selection). An illustrative case is the *invected* gene of *D. melanogaster* (also listed in GadFly as CG17835). This gene encodes a homeodomain protein that is similar to *engrailed*, and these two genes are adjacent. One of four *invected* exons is only 6 nucleotides long and is flanked by introns of 27,659 and 1,134 nucleotides. Significantly, this exon is not recognized by cDNA alignment software such as SIM4 (Florea et al. 1998), and the gene is incorrectly annotated (GenBank accession number AE003825.1). As a result, the protein sequence predicted by annotation of the genome (Adams et al. 2000; GenBank accession number AAF58640) differs from that predicted from the cDNA (Coleman et al. 1987; GenBank accession number CAA28885), because of a frameshift affecting the entire carboxyl-terminal coding exon, a highly conserved region of the protein. This is despite the fact that the microexon sequence, GTCGAA, is flanked by intron sequences that perfectly match the splice-site consensus. Use of this microexon provides perfect agreement between the cDNA and genomic sequences when consensus splice sites are used, whereas the annotation predicts an RNA with several discrepancies relative to the cDNA. The frameshift is due to the predicted use of a 5' splice site 10 nucleotides downstream of the true 5' splice site, which was apparently selected to account for the microexon. It seems clear that the protein sequence predicted by the cDNA is correct. Why was it not incorporated into the annotation? The alignment problem

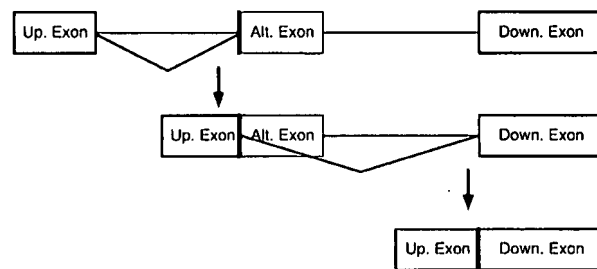
arises because a pattern-matching algorithm that locates exons by similarity between the cDNA and the genomic sequence cannot find exons of this size unless its stringency is reduced to an unacceptable level (Florea et al. 1998).

The notion that exons can be arbitrarily small is supported by the observation of exons with length 0. Of course, such sites are not exons at all but, rather, are resplicing sites (see fig. 1). This phenomenon has been demonstrated in the case of the *Drosophila Ultrathorax* locus (Hatton et al. 1998), which has a region of 60 kb containing two alternatively spliced exons, and may be a general feature of long introns (J. Burnette and A. J. Lopez, personal communication). The existence of resplicing sites not only illustrates the lack of a lower limit to exon size (which has implications for gene annotation) but also has implications for the analysis of hereditary mutations. A mutation at one of these sites could potentially create a frozen intermediate such as that diagrammed in figure 1. This partially spliced RNA would probably be unstable, because of nonsense-mediated decay (Culbertson 1999), and the apparent result would be no RNA (rather than aberrantly spliced RNA). Such mutations would be very hard to identify.

#### Nucleotides Far from Splice Sites Can Affect Splicing

No method of evaluating potential splice sites that is based on sequence alone can be 100% reliable. One can be sure of this because many sequences that are not splice sites are capable of acting as splice sites, and vice versa. Perhaps the clearest demonstration of this is provided by the activation of cryptic splice sites. These are splice sites that are used, sometimes with 100% efficiency, when a natural splice site has been mutationally inactivated. The activation of cryptic sites occurs in approximately one-third of splicing mutations (Nakai and Sakamoto 1994). The phenomenon shows that the cryptic sites are perfectly capable of being recognized by the splicing machinery. Clearly, the sequence of such cryptic sites is compatible with splicing, and context is important for splice-site choice.

Two contextual elements that contribute to splice-site selection are the location of splice sites relative to each other and splicing-enhancer sequences. The exon-size preferences described above are widely understood in terms of an exon-definition model that includes the interaction of splicing factors bound at either end of an exon (Berget 1995). The requirement for productive interactions among splicing factors, including U1 snRNPs at the 5' splice site and U2 snRNP auxiliary factor (U2AF) at the 3' splice site, are thought to give rise to preferred exon lengths because of steric constraints and geometry favoring interactions. In the case of small introns, a similar model of intron bridging has been pro-



**Figure 1** Small internal exons and resplicing. This schematic figure indicates the pathway of resplicing demonstrated for the *Drosophila Ubx* locus (Hatton et al. 1998). The thicker vertical line indicates a resplicing site, which does not contribute any nucleotides to the final mRNA product. The same pathway could be followed in the case of a microexon, in which case an arbitrarily small number of nucleotides would remain in the mRNA product. "Up. Exon" and "Down. Exon" denote the exons upstream and downstream of the resplicing site, respectively. In the case of *Ubx*, the sequence immediately downstream of the resplicing site is an alternatively spliced exon (here designated "Alt. Exon"), but resplicing sites are not always accompanied by such alternatively spliced exons (J. Burnette and A. J. Lopez, personal communication).

posed (Guo and Mount 1995; McCullough and Berget 1997). In combination, these models suggest that, in order to be recognized, a splice site must have a partner an appropriate distance away, so that either exon definition or intron definition is facilitated by the spacing. One experimental distinction between exon definition and intron definition is the result of mutations that inactivate the splice site. Failure to undergo exon definition results in exon skipping, whereas failure to undergo intron definition results in intron retention.

Not only is the use of one splice site dependent on the presence of its partner across the exon, but weakness in one partner can be compensated by strength in the other, as seen with second-site revertants of splice-site mutations that cause exon skipping. In an analysis of splicing mutations at the dihydrofolate reductase locus, Carothers et al. (1993) found that a mutation at the 5' splice site of exon 5 (G to C in the third position of the intron) could be partially reversed by mutations that increased the strength of the 3' splice site upstream of the same exon (AAAG| to TTAG|, ACAG|, or ATAG|). Although reversion was not complete, these data provide a strong argument that whether a sequence functions as a splice site depends not only on its intrinsic strength but also on its context. Similarly, there are mutations that create splice sites within introns, activating cryptic exons by recruitment of appropriately placed partners (e.g., see Bagnall et al. 1999).

Splicing enhancers are sequences that stimulate splicing at nearby sites. A family of non-snRNP splicing factors known as "SR proteins" appear to be important for the recognition of splicing enhancers in

exons (Blencowe 2000). A splicing difference between SMN1 and SMN2, which explains their differential effects on spinal muscular atrophy, has been attributed to a translationally silent substitution within the coding sequence that affects splicing (Lorson et al. 1999). Similarly, H.-X. Liu, L. Cartegni, M. Q. Zhang, and A. R. Krainer (personal communication) have shown that a nonsense mutation causing the skipping of BRCA1 exon 18 affects splicing in vitro and that a missense mutation at the same position can also cause exon skipping. There are also splicing-enhancer sequences in introns—and examples of mutations that affect them (Cogan et al. 1997). Although general mechanisms for their function have yet to be defined, there is some evidence that at least some splicing enhancers in introns may act by facilitating exon definition in the case of small exons (Carlo et al. 2000).

### Outlook

This review has presented aspects of pre-mRNA splicing that pose special problems for gene annotation. However, even though the best gene finders predict genes exactly right less than half the time, 95% of total coding nucleotides are predicted accurately, and <5% of genes are completely missed (Reese et al. 2000; Genome Annotation Assessment Project-GASP1). When cDNA and homology data are available, annotations will tend to be even better. Thus, one would be wrong to conclude from this review that the gene annotations attending the human genome sequence will not provide an extremely valuable resource. Nevertheless, molecular geneticists will want to have an understanding of the kinds of errors that are likely to occur—and to carefully review the available evidence for genes that matter to them. Annotators are likewise obligated to make the source of each specific aspect of their annotation an integral part of the annotation; for example, if part of the annotation is supported by a EST whereas the rest of it is based on the prediction of a gene finder, then the limits of the cDNA should be indicated, and the accession number of the EST should be part of the annotation.

A related but distinct point is that these same factors are also relevant when candidate mutations are evaluated during the analysis of hereditary disease. Mutations that lie within splicing enhancers, at resplicing sites, or at cryptic splice sites can affect splicing even when they lie some distance from the splice sites actually used in the generation of the affected mRNA. The problem is further compounded by alternative splicing and the interplay between splicing and polyadenylation, topics that are beyond the scope of the present review.

In summary, gene annotations will be a valuable resource. However, they will not substitute for expertise in molecular genetics.

### Acknowledgments

Support by National Institutes of Health grant GM37991-11 is gratefully acknowledged. I thank Doug Black for helpful comments on the manuscript. I thank James Burnette, A. Javier Lopez, and Adrian Krainer for providing information prior to publication.

### Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Computational Genefinding, <http://www.cse.ucsc.edu/~haussler/genefindingpaper>  
 GadFly: Genome Annotation Database of Drosophila, <http://www.fruitfly.org/annot/index.html>  
 GenBank, <http://www.ncbi.nlm.nih.gov/> (for incorrect annotation of *invected* [accession number AE003825.1] and predicted protein sequence [accession numbers AAF58640 and CAA28885])  
 Gene Annotation and Splice Site Selection, <http://www.wam.umd.edu/~smount/Annotation.html>  
 Genome Annotation Assessment Project-GASP1, <http://www.fruitfly.org/GASP1/index.html>

### References

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
- Aebi M, Hornig H, Padgett RA, Reiser J, Weissman C (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47:555–565
- Bagnall RD, Waseem NH, Green PM, Colvin B, Lee C, Giannelli F (1999) Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. *Br J Haematol* 107:766–771
- Berget SM (1995) Exon recognition in vertebrate splicing. *J Biol Chem* 270:2411–2414
- Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25:106–110
- Bork P (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* 10:398–400
- Burge CB, Padgett RA, Sharp PA (1998) Evolutionary fates and origins of U12-type introns. *Mol Cell* 2:773–785
- Carlo T, Sierra R, Berget SM (2000) A 5' splice site-proximal enhancer binds SF1 and activates exon bridging. *Mol Cell Biol* 20:3988–3995
- Carothers AM, Urlaub G, Grunberger D, Chasin LA (1993) Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol Cell Biol* 13:5085–5098
- Cogan JD, Prince MA, Lekhakula S, Bundrey S, Futrakul A, McCarthy EM, Phillips JA III (1997) A novel mechanism of aberrant pre-mRNA splicing in humans. *Hum Mol Genet* 6:909–912
- Coleman KG, Poole SJ, Weir MP, Soeller WC, Kornberg T

- (1987) The invected gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the engrailed gene. *Genes Dev* 1:19-28
- Culbertson MR (1999) RNA surveillance: unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends Genet* 15:74-80
- Dietrich RC, Incurvaia R, Padgett RA (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell* 1:151-160
- Dominski Z, Kole R (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* 11:6075-6083
- Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA. *Genome Res* 8:967-974
- Gilbert W (1978) Why genes in pieces? *Nature* 271:501
- Guo M, Mount SM (1995) Localization of sequences required for size-specific splicing of a small *Drosophila* intron. *J Mol Biol* 253:426-437
- Hall SL, Padgett RA (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol* 239:357-365
- Hatton AR, Subramaniam V, Lopez AJ (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol Cell* 2:787-796
- Hawkins JD (1988) A survey on intron and exon lengths. *Nucleic Acids Res* 16:9893-9905
- Jackson IJ (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 19:3795-3798
- Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci USA* 96:6307-6311
- Maroni G (1996) The organization of eukaryotic genes. *Evol Biol* 29:1-19
- McCullough AJ, Berget SM (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol* 17:4562-4571
- Mount SM (1982) A catalogue of splice junction sequences. *Nucleic Acids Res* 10:459-472
- Nakai K, Sakamoto H (1994) Construction of a novel database containing aberrant splicing mutations of mammalian gene. *Gene* 141:171-177
- Reese MG, Hartzell G, Harris NL, Ohler U, Abril JF, Lewis SE (2000) Genome annotation assessment in *Drosophila melanogaster*. *Genome Res* 10:483-501
- Robberson BL, Cote GL, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10:84-94
- Roller AB, Hoffman DC, Zahler AM (2000) The allele-specific suppressor sup-39 alters use of cryptic splice sites in *Caenorhabditis elegans*. *Genetics* 154:1169-1179
- Senapathy P, Sharpiro MB, Harris NL (1990) Splice junctions, branch point sites and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183:252-278
- Shukla GC, Padgett RA (1999) Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* 5:525-538
- Sidrauski C, Cox JS, Walter P (1996) tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. *Cell* 87:405-413
- Stamm S, Zhang MQ, Marr TG, Helfman DM (1994) A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res* 22:1515-1526
- Stormo GD (2000) Gene-finding approaches for eukaryotes. *Genome Res* 10:394-397
- Tarn WY, Steitz JA (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* 22:132-137
- Wu A, Krainer AR (1999) AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol Cell Biol* 19:3225-3236
- Zhang MQ (1998) Statistical features of human exons and their flanking regions. *Hum Mol Genet* 7:919-932

---

**Codon usage in plant genes**

---

---

Elizabeth E. Murray<sup>1,2\*</sup>, Jeff Lotzer<sup>1</sup> and Mary Eberle<sup>1</sup>

---

<sup>1</sup>Agrigenetics Advanced Sciences Company, 5649 East Buckeye Road, Madison, WI 53713 and<sup>2</sup>McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI 53706, USA

---

Received October 6, 1988; Accepted December 14, 1988

---

**Abstract:**

We have examined codon bias in 207 plant gene sequences collected from Genbank and the literature. When this sample was further divided into 53 monocot and 154 dicot genes, the pattern of relative use of synonymous codons was shown to differ between these taxonomic groups, primarily in the use of G+C in the degenerate third base. Maize and soybean codon bias were examined separately and followed the monocot and dicot codon usage patterns respectively. Codon preference in ribulose 1,5 biphosphate and chlorophyll a/b binding protein, two of the most abundant proteins in leaves was investigated. These highly expressed are more restricted in their codon usage than plant genes in general.

**Introduction:**

With the exception of Met and Trp, all amino acids are encoded by two to six synonymous codons. In the majority of species studied to date, an organism's use of synonymous codons is not random (1-5). However, detailed characterization of specific patterns of codon usage have been reported primarily for unicellular organisms, including *E. coli* (6-8), *Bacillus* (9,10), *Agrobacterium* (11) and yeast (12-17). The pattern of codon usage in higher eukaryotes has been examined in only a limited number of species, including *Drosophila* (19) and man (13, 18-21).

In the last three years, a large number of DNA sequences of higher plant genes have been reported, enabling us to extend the initial analyses of plant codon usage previously reported (13, 18, 22,23). We have used an expanded sample of 207 plant gene sequences to examine some general observations about codon usage.

In general, genes within a taxonomic group exhibit similarities in codon choice, regardless of the function of these genes. Thus an estimate of the overall use of the genetic code by an taxonomic group can be obtained by summing codon frequencies of all its sequenced genes. This species-specific codon choice has been called a "codon dialect" by Ikemura (13). Here we report on the "codon dialect" of 207 plant genes. We have broken this sample down into monocotyledonous and dicotyledonous plants to determine whether these broader taxonomic groups are characterized by different patterns of synonymous codon preference. Finally, we report the codon dialect of maize and soybean, since over 25 genes have been sequenced in each of these agronomically important species.

Bias in codon choice within genes in a single species appears related to the level of expression of the protein encoded by that gene (6, 7, 12-18). Codon bias is most extreme in highly expressed proteins of *E. coli* and yeast. In these organisms, a strong positive

correlation has been reported between the abundance of an isoaccepting tRNA species and the favored synonymous codon. In one group of highly expressed proteins in yeast, over 96% of the amino acids are encoded by only 25 of the 61 available codons (15). These 25 codons are preferred in all sequenced yeast genes, but the degree of preference varies with the level of expression of the genes. Recently, Hoekema and colleagues (24) report that replacement of these 25 preferred codons by minor codons in the 5' end of the highly expressed yeast gene PGK1 results in a decreased level of both protein and mRNA. They conclude that biased codon choice in highly expressed genes enhances translation and is required for maintaining mRNA stability in yeast. The degree of codon bias may be a factor to consider when engineering high expression of heterologous genes in yeast and other systems.

In plants, the ribulose 1,5 biphosphate small subunit (RuBPC SSU) and chlorophyll a/b binding protein (CAB) gene families encode two of the most abundant proteins in leaves. These genes have been sequenced in a number of different plants, and we have examined their codon usage in detail to determine whether they are more biased than other plant genes.

### **Data and Methods:**

The 207 plant genes included in the sample are detailed in Tables 1 and 2. Sources for the data were Genbank (release 55) or original publications, referenced in Tables 1 and 2 and listed in the Appendix. Genes for which only genomic sequence was available were included, and introns were removed before codon usage data was generated. Partial sequences available for some genes were included when reading frame could be determined.

Homologous genes in different species, or multigene gene families within a species have been sequenced, including those for zein, glycinin, vicilin, CAB and RuBPC SSU. Multiple sequences of these genes were included in the sample if they differed by a minimum of 10% in the base composition of their protein coding regions. As a result, this sample may contain some bias towards codon usage in highly expressed genes.

Genbank sequences were extracted using the GENBANK program (25). Codon usage tables were compiled using the program CODONFREQUENCY from the program library of the University of Wisconsin Genetics Computer group (26).

### **Results:**

Plant genes coding for proteins with a wide variety of functions have now been sequenced. We have tabulated the sequences of 207 plant genes (Tables 1 and 2) from 6 monocot and 36 dicot species. These proteins are present in a wide range of plant tissues at varying levels of expression. However, to date, only a few plant genes encoding rare proteins and/or mRNAs have been sequenced.

We have calculated the codon usage profile of the pooled plant sample and separate codon usage profiles for the monocotyledonous and dicotyledonous groups of plants (Table 3). This division reveals that the relative use of synonymous codons differs between the monocots and the dicots. Since the monocot sample is one third smaller than the dicot sample, we were concerned that the relative abundance of storage protein genes could skew its codon usage profile. Accordingly, we calculated codon usage in the monocots without these genes (Table 3).

Table 1  
Descriptions and sources of 53 monocot genes included in the analysis.

GENUS/SPECIES	GENBANK	PROTEIN	REF
<i>Avena sativa</i>	ASTAP3R	Phytochrome 3	1
<i>Hordeum vulgare</i>	BLYALR	Aleurain	2
	BLYAMY1	$\alpha$ amylase 1	3
	BLYAMY2	$\alpha$ amylase 2	3
	BLYCHORD1	Hordein C	4
	BLYGLUCB	$\beta$ glucanase	5
	BLYHORB	B1 hordein	6
	BLYPAP1	Amylase/protease inhibitor	7
	BLYTHIAR	Toxin $\alpha$ hordothionin	8
	BLYUBIQR	Ubiquitin	9
		Histone 3	10
		Leaf specific thionin 1	11
		Leaf specific thionin 2	11
		Plastocyanin	12
<i>Oryza sativa</i>	RICGLUTG	Glutelin	13
		Glutelin	14
<i>Triticum aestivum</i>	WHTAMYA	$\alpha$ amylase	15
	WHTCAB	CAB	16
	WHTEMR	Em protein	17
	WHTGIR	gibberellin responsive protein	18
	WHTGLGB	$\gamma$ gliadin	19
	WHTGLIABA	$\alpha/\beta$ gliadin Class AII	20
	WHTGLUT1	High MW glutenin	21
	WHTH3	Histone 3	22
	WHTH4091	Histone 4	23
	WHTRBCB	RuBPC small subunit	24
<i>Secale cereale</i>	KYESECGSR	$\gamma$ secalin	25
<i>Zea mays</i>	MZEA1G	40.1 kD A1 protein (NADPH-dependent reductase)	26
	MZEACT1G	Actin	27
	MZEADH11F	Alcohol dehydrogenase 1	28
	MZEADH2NR	Alcohol dehydrogenase 2	28
	MZEALD	Aldolase	29
	MZEANT	ATP/ADP translocator	30
	MZEEG2R	Glutelin 2	31
	MZEGGST3B	Glutathione S transferase	32
	MZEH3C2	Histone 3	33
	MZEH4C14	Histone 4	34
	MZEHSP701	70 kD Heat shock protein, exon 1	35
	MZEHSP702	70 kD Heat shock protein, exon 2	35
	MZELHCP	CAB	36
	MZEMPL3	Lipid body surface protein L3	37
	MZEPEPCR	Phosphoenolpyruvate carboxylase	38
	MZERBCS	RuBPC small subunit	39
	MZESUSYSG	Sucrose synthetase	40
	MZETP12	Triosephosphate isomerase 1	41
	MZEZE20M	19 kD zein	42
	MZEZE30M	19 kD zein	42
	MZEZE15A3	15 kD zein	43
	MZEZE16	16 kD zein	44
	MZEZE19A	19 kD zein	45
	MZEZE22A	22 kD zein	46
	MZEZE22B	22 kD zein	46
		Catalase 2	47
		Regulatory C1 locus	48

Data was obtained from GenBank (release 55) or, when no Genbank file name is specified, directly from the published source.

In general, the most important factor in discriminating between monocot and dicot patterns of codon usage is the percentage G+C content of the degenerate third base. In monocots, 16 of 18 amino acids favor G+C in this position, while dicots only favor G+C in 7 of 18 amino acids.



Table 2  
Descriptions and sources of 154 genes dicot included in the analysis.

GENUS/ SPECIES	GENBANK	PROTEIN	REF
<i>Anthriscum majus</i>	AMACHS	Chalcone synthetase	49
<i>Arabidopsis thaliana</i>	ATHADH	Alcohol dehydrogenase	50
	ATHH3GA	Histone 3 gene 1	51
	ATHH3GB	Histone 3 gene 2	51
	ATHH4GA	Histone 4 gene 1	51
	ATHLHCP1	CAB	52
	ATHTUBA	$\alpha$ tubulin	53
		5-enolpyruvyl4hifate 3-phosphate synthetase	54
<i>Bertholletia excelsa</i>		High methionine storage protein	55
<i>Brassica campestris</i>		Acyl carrier protein	56
<i>Brassica napus</i>	BNANAP	Napin	57
<i>Brassica oleracea</i>	BOLSLSGR	S-locus specific glycoprotein	58
<i>Canavalia ensiformis</i>	CENCONA	Concanavalin A	59
<i>Carica papaya</i>	CPAPAP	Papain	60
<i>Chlamydomonas reinhardtii</i>	CRECS2	Preapocytochrome	61
	CRERBCS1	RuBPC small subunit gene 1	62
	CRERBCS2	RuBPC small subunit gene 2	62
<i>Cucurbita pepo</i>	CUCPHT	Phytochrome	63
<i>Cucumis sativus</i>	CUSGMS	Glyoxosomal malate synthetase	64
	CUSLHCPA	CAB	65
	CUSSSU	RuBPC small subunit	65
<i>Daucus carota</i>	DAREXT	Extensin	66
	DAREXTR	33 kD extensin related protein	67
<i>Dolichos biflorus</i>	DBILECS	seed lectin	68
<i>Flaveria trinervia</i>	FTRBCR	RuBPC small subunit	69
<i>Glycine max</i>	SOY7SAA	7S storage protein	70
	SOYACT1G	Actin 1	27
	SOYCIPI	CII protease inhibitor	71
	SOYGLYA1A	Glycinin A1a Bx subunits	72
	SOYGLYAA8	Glycinin A5A4B3 subunits	73
	SOYGLYAB	Glycinin A3/b4 subunits	74
	SOYGLYR	Glycinin A2B1a subunits	75
	SOYHSP175	Low M W heat shock proteins	76
	SOYLGBI	Leghemoglobin	77
	SOYLEA	Lectin	78
	SOYLOX	Lipoxygenase 1	79
	SOYNOD20G	20 kDa nodulin	80
	SOYNOD23G	23 kDa nodulin	81
	SOYNOD24H	24 kDa nodulin	82
	SOYNOD26B	26 kDa nodulin	83
	SOYNOD26R	26 kDa nodulin	84
	SOYNOD27R	27 kDa nodulin	83
	SOYNOD35M	35 kDa nodulin	86
	SOYNOD75	75 kDa nodulin	86
	SOYNODR1	Nodulin C51	87
	SOYNODR2	Nodulin E27	87
	SOYPRP1	Proline rich protein	88
	SOYRUBP	RuBPC small subunit	89
	SOYURA	Urease	90
	SOYHSP26A	Heat shock protein 26A	91
		Nuclear-encoded chloroplast heat shock protein	92
		22 kDa nodulin	80
		$\beta$ 1 tubulin	93
		$\beta$ 2 tubulin	93
<i>Gossypium hirsutum</i>		Seed $\alpha$ globulin (vicilin)	94
		Seed $\beta$ globulin (vicilin)	94
<i>Helianthus annuus</i>	HNNRUBCS	RuBPC small subunit	95
		2S albumin seed storage protein	96
<i>Ipomoea batatas</i>		Wound-induced catalase	97
<i>Lemna gibba</i>	LGIAB19	CAB	98
	LGIR5BPC	RuBPC small subunit	99
<i>Lupinus luteus</i>	LUPLBR	leghemoglobin I	100

GENUS/ SPECIES	GENBANK	PROTEIN	REF
<i>Lycopersicon esculentum</i>	TOMBIOBR	Biotin binding protein	101
	TOMETHYBR	Ethylene biosynthesis protein	102
	TOMPG2AR	Polygalacturonase-2a	103
	TOMPSI	Tomato photosystem I protein	104
	TOMRBCSA	RuBPC small subunit	105
	TOMRBCSB	RuBPC small subunit	105
	TOMRBCSC	RuBPC small subunit	105
	TOMRBCSD	RuBPC small subunit	106
	TOMRRD	Ripening related protein	107
	TOMWIPIG	Wound induced proteinase inhibitor I	108
	TOMWIPII	Wound induced proteinase inhibitor II	109
		CAB 1A	110
		CAB 1B	110
		CAB 3C	110
		CAB 4	111
		CAB 5	111
		Leghemoglobin III	112
<i>Medicago sativa</i>	ALFLB3R		
<i>Mesembryanthemum crystallinum</i>		RuBPC small subunit	113
<i>Nicotiana plumbaginifolia</i>	TOBATP21	Mitochondrial ATP synthase $\beta$ subunit	114
		Nitrate reductase	115
		Glutamine synthetase	116
<i>Nicotiana tabacum</i>	TOBECH	Endochitinase	117
	TOBGAPA	A subunit of chloroplast G3PD	118
	TOBGAPB	B subunit of chloroplast G3PD	118
	TOBGAPC	C subunit of chloroplast G3PD	118
	TOBPRIAR	Pathogenesis related protein 1a	119
	TOBPRICR	Pathogenesis-related protein 1c	119
	TOBPRPR	Pathogenesis related protein 1b	120
	TOBPXDLF	Peroxidase	121
	TOBRBPCO	RuBPC small subunit	122
	TOBTHAUR	TMV-induced protein homologous to thaumatin	123
<i>Persea americana</i>	AVOCEL	Cellulase	124
<i>Petroselinum hortense</i>	PHOCHL	Chalcone synthase	125
<i>Petunia sp.</i>	PETCAB13	CAB 13	126
	PETCAB22L	CAB 22L	126
	PETCAB22R	CAB 22R	126
	PETCAB25	CAB 25	126
	PETCAB37	CAB 37	127
	PETCAB91R	CAB 91R	127
	PETCHSR	Chalcone synthase	128
	PETGCR1	Glycine-rich protein	129
	PETRBCS08	RuBPC small subunit	130
	PETRBCS11	RuBPC small subunit	130
		70 kDa heat shock protein	131
<i>Phaseolus vulgaris</i>	PHVCHM	Chitinase	132
	PHVDLECA	Phytohemagglutinin E	133
	PHVDLECB	Phytohemagglutinin L	133
	PHVGSR1	Glutamine synthetase 1	134
	PHVGSR2	Glutamine synthetase 2	134
	PHVLBA	Leghemoglobin	135
	PHVLECT	Lectin	136
	PHVPAL	Phenylalanine ammonia lyase	137
	PHVPHASAR	$\alpha$ phaseolin	138
	PHVPHASBR	$\beta$ phaseolin	138
		Arcein seed protein	139
<i>Pisum sativum</i>	PEAALB2	Seed albumin	140
	PEACAB80	CAB	141
	PEAGSR1	Glutamine synthetase (nodule)	142

GENUS/ SPECIES	GENBANK	PROTEIN	REF
	PEALECA	Lectin	144
	PEALEGA	Legumin	145
	PEARUBPS	RuBPC small subunit	146
	PEAVIC2	Vicilin	147
	PEAVIC4	Vicilin	147
	PEAVIC7	Vicilin	147
		Alcohol dehydrogenase 1	148
		Glutamine synthetase (leaf)	143
		Glutamine synthetase (root)	143
		Histone 1	149
		Nuclear encoded chloroplast heat shock protein	92
<i>Raphanus sativus</i>		RuBPC small subunit	150
<i>Ricinus communis</i>	RCCAGG	Agglutinin	151
	RCCRICIN	Ricin	152
	RCCIC14	Isocitrate lyase	153
<i>Silene pratensis</i>	SIPFDX	Ferredoxin precursor	154
	SIPPCY	Plastocyanin precursor	155
<i>Sinapis alba</i>	SALGAPDH	Nuclear gene for G3PD	156
<i>Solanum tuberosum</i>	POTPAT	Patatin	157
	POTINHWI	Wound-induced proteinase inhibitor	158
	POTLS1G	Light-inducible tissue specific ST-LS1 gene	159
	POTPI2G	Wound-induced proteinase inhibitor II	160
	POTRBCS	RuBPC small subunit	161
		Sucrose synthetase	162
<i>Spinacia oleracea</i>	SPIACP1	Acyl carrier protein I	163
	SPIOEC16	16 kDa photosynthetic oxygen-evolving protein	164
	SPIOEC23	23 kDa photosynthetic oxygen-evolving protein	165
	SPIPCG	Plastocyanin	165
	SPIPS33	33 kDa photosynthetic water oxidation complex precursor	166
		Glycolate oxidase	167
<i>Vicia faba</i>	VFALBA	Leghemoglobin	168
	VFALEB4	Legumin B	169
		Vicilin	170

Data was obtained from GenBank (release 55) or, when no Genbank file name is specified, directly from the published source.

Table 3  
Codon usage in pooled sequences of higher plant genes.

AmAcid	Codon	Plants n = 207		Dicots n = 154		Monocots n = 53		Monocots No Storage Proteins n = 39	
		No.	%	No.	%	No.	%	No.	%
Gly	GGG	731	15	449	12	282	21	267	22
Gly	GGA	1629	32	1399	38	230	17	193	16
Gly	GGT	1477	29	1231	34	246	18	207	17
Gly	GGC	1179	24	596	16	583	44	543	45
Glu	GAG	2102	57	1498	51	604	75	568	79
Glu	GAA	1616	43	1419	49	197	25	154	21
Asp	GAT	1458	50	1269	58	189	27	162	24
Asp	GAC	1441	50	927	42	514	73	503	76
Val	GTG	1354	31	956	29	398	36	338	37
Val	GTA	491	11	402	12	89	8	52	6
Val	GTT	1478	34	1270	39	208	19	154	17
Val	GTC	1045	24	642	20	403	37	362	40

AmAcid	Codon	Plants		Dicots		Monocots		Monocots No Storage Proteins	
		No.	%	No.	%	No.	%	No.	%
Ala	GCG	546	11	211	6	335	22	284	24
Ala	GCA	1156	22	916	25	240	16	137	12
Ala	GCT	1901	37	1530	42	371	24	254	21
Ala	GCC	1548	30	960	27	588	38	510	43
Arg	AGG	742	26	540	25	202	26	163	25
Arg	AGA	707	24	633	30	74	9	50	8
Ser	AGT	581	13	493	14	88	8	51	6
Ser	AGC	887	20	605	18	282	26	225	27
Lys	AAG	2241	66	1600	61	641	86	609	86
Lys	AAA	1139	34	1034	39	105	14	98	14
Asn	AAT	1137	41	982	45	155	25	106	23
Asn	AAC	1646	59	1188	55	458	75	356	77
Met	ATG	1356	100	982	100	374	100	311	100
Ile	ATA	505	16	419	18	86	11	48	8
Ile	ATT	1241	40	1051	45	190	24	128	21
Ile	ATC	1374	44	873	37	501	65	433	71
Thr	ACG	343	11	184	8	159	21	146	22
Thr	ACA	745	24	636	27	109	14	73	11
Thr	ACT	990	31	842	35	148	19	116	18
Thr	ACC	1082	34	721	30	361	46	319	49
Trp	TGG	790	100	605	100	185	100	171	100
End	TGA	68	33	50	33	18	34	15	37
Cys	TGT	432	40	338	44	94	30	69	27
Cys	TGC	647	60	423	56	224	70	185	73
End	TAG	48	24	29	19	19	36	11	27
End	TAA	88	43	72	48	16	30	14	35
Tyr	TAT	743	37	630	43	113	21	68	16
Tyr	TAC	1267	63	838	57	429	79	354	84
Leu	TTG	1185	22	965	26	220	14	108	9
Leu	TTA	412	8	363	10	49	3	19	2
Phe	TTT	1047	40	887	45	160	25	95	20
Phe	TTC	1597	60	1106	55	491	75	392	80
Ser	TCG	343	8	192	6	151	14	139	17
Ser	TCA	768	17	649	19	119	11	67	8
Ser	TCT	1009	22	844	25	165	15	112	13
Ser	TCC	896	20	621	18	275	26	237	29
Arg	CGG	198	7	95	5	103	13	94	14
Arg	CGA	214	7	181	8	33	4	23	3
Arg	CGT	534	18	441	21	93	12	67	10
Arg	CGC	520	18	241	11	279	36	268	40
Gln	CAG	1465	43	787	41	678	46	457	60
Gln	CAA	1912	57	1125	59	787	54	305	40
His	CAT	575	48	465	54	110	33	85	30
His	CAC	625	52	398	46	227	67	202	70
Leu	CTG	792	15	347	9	445	28	371	33
Leu	CTA	434	8	281	8	153	9	59	5
Leu	CTT	1273	24	1032	28	241	15	151	13
Leu	CTC	1189	22	691	19	498	31	434	38
Pro	CCG	492	13	236	9	256	23	224	30
Pro	CCA	1507	39	1126	42	381	34	202	27
Pro	CCT	1063	28	874	32	189	17	118	15
Pro	CCC	755	20	469	17	286	26	212	28

n = the number of DNA sequences in the sample. No. is the number occurrences of a given codon in the sample. % is the percent occurrence for each codon within a given amino acid in the sample. (See text for description of the samples).

Table 4  
Codon usage in pooled sequences of higher plant genes.

AmAcid	Codon	Soybean n = 29		Maize n = 26		CAB n = 17		RuBP SSU n = 20	
		No.	%	No.	%	No.	%	No.	%
Gly	GGG	90	16	95	16	42	8	16	9
Gly	GGA	189	33	78	13	167	32	95	51
Gly	GGT	193	33	129	21	196	37	32	17
Gly	GGC	102	18	302	50	118	23	43	23
Glu	GAG	310	51	368	81	178	71	139	74
Glu	GAA	301	49	84	19	73	29	49	26
Asp	GAT	244	62	87	24	53	29	39	33
Asp	GAC	150	38	277	76	128	71	79	67
Val	GTG	219	37	227	40	62	21	93	36
Val	GTA	77	13	36	6	24	8	7	3
Val	GTT	227	38	99	17	118	39	87	33
Val	GTC	75	12	209	37	96	32	73	28
Ala	GCG	42	8	211	24	26	5	16	5
Ala	GCA	170	30	115	13	61	12	42	14
Ala	GCT	208	37	237	27	225	45	110	38
Ala	GCC	139	25	324	36	192	38	125	43
Arg	AGG	88	22	109	26	21	15	17	12
Arg	AGA	119	30	28	7	33	24	31	21
Ser	AGT	117	18	29	5	15	5	21	8
Ser	AGC	129	20	150	28	84	27	56	22
Lys	AAG	278	58	367	90	186	85	176	85
Lys	AAA	204	42	43	10	34	15	30	15
Asn	AAT	168	40	56	19	52	30	35	26
Asn	AAC	248	60	246	81	119	70	102	74
Met	ATG	184	100	210	100	111	100	115	100
Ile	ATA	109	24	35	8	10	6	1	1
Ile	ATT	219	49	100	24	61	40	63	43
Ile	ATC	118	27	284	68	83	54	83	56
Thr	ACG	29	7	114	26	10	6	5	3
Thr	ACA	128	29	48	11	35	22	21	13
Thr	ACT	151	35	72	16	61	38	59	36
Thr	ACC	124	29	212	47	54	34	79	48
Trp	TGG	82	100	84	100	99	100	86	100
End	TGA	5	18	7	26	15	88	2	11
Cys	TGT	63	40	29	21	16	39	7	9
Cys	TGC	95	60	110	79	25	61	72	91
End	TAG	9	32	14	52	0	0	1	5
End	TAA	14	50	6	22	2	12	16	84
Tyr	TAT	135	49	38	14	23	19	17	10
Tyr	TAC	139	51	240	86	99	81	151	90
Leu	TTG	175	24	116	13	118	30	79	36
Leu	TTA	79	11	28	3	15	4	6	3
Phe	TTT	166	46	69	20	106	40	32	20
Phe	TTC	193	54	278	80	160	60	125	80
Ser	TCG	39	6	89	16	17	5	10	4
Ser	TCA	125	19	56	10	46	15	48	19
Ser	TCT	140	22	75	14	83	26	33	13
Ser	TCC	94	15	145	27	69	22	89	34
Arg	CGG	17	4	54	13	7	5	1	1
Arg	CGA	41	10	13	3	6	4	3	2
Arg	CGT	70	18	45	11	50	36	48	33
Arg	CGC	64	16	165	40	20	15	44	31
Gln	CAG	181	41	311	59	36	37	75	51
Gln	CAA	261	59	219	41	60	62	73	49
His	CAT	124	63	49	29	16	32	4	18
His	CAC	73	37	122	71	34	68	18	82

AmAcid	Codon	Soybean		Maize		CAB		RuBP SSU	
		No.	%	No.	%	No.	%	No.	%
Leu	CTG	75	10	289	31	29	7	27	12
Leu	CTA	60	8	78	9	6	2	9	4
Leu	CTT	184	26	147	16	134	34	56	25
Leu	CTC	148	21	261	28	88	23	43	20
Pro	CCG	55	8	149	27	29	10	13	6
Pro	CCA	346	47	126	23	137	47	72	34
Pro	CCT	236	32	109	20	73	25	60	29
Pro	CCC	95	13	164	30	54	18	66	31

$n$  = the number of DNA sequences in the sample. No. is the number occurrences of a given codon in the sample. % is the percent occurrence for each codon within a given amino acid in the sample. (See text for description of the samples).

The G ending codons for Thr, Pro, Ala and Ser are avoided in both monocots and dicots because they contain C in codon position II. The CG dinucleotide is strongly avoided in plants (23) and other eukaryotes (27), possibly due to regulation involving methylation. In dicots, XCG is always the least favored codon, while in monocots this is not the case. The doublet TA is also avoided in codon positions II and III in most eukaryotes (27), and this is true of both monocots and dicots.

Grantham and colleagues (18) have developed two codon choice indices to quantify CG and TA doublet avoidance in codon positions II and III. XCG/XCC is the ratio of codons having C as base II of G-ending to C-ending triplets, while XTA/XTT is the ratio of A-ending to T-ending triplets with T as the second base. These indices have been calculated for the plant data in this paper (Table 5) and support the conclusion that monocot and dicot species differ in their use of these dinucleotides. This pattern of synonymous codon usage is not dependent on the inclusion of storage protein genes in the monocot sample, since the pooled codon usage data for monocots without storage proteins is even less like the dicot pattern (Table 3). Not surprisingly, the pooled plant sample resembles the dicot pattern more than that of monocots, since almost three times as many dicot sequences as monocot sequences were available.

For two species, soybean and maize, larger sequence samples were available (29 and 28 genes respectively), so that species-specific codon usage profiles could be calculated (Table 4). Not surprisingly, the maize codon usage pattern resembles that of monocots in general, since these sequences represent over half of the monocot sequences available. The codon profile of the maize subsample is even more strikingly biased in its preference for G+C in

Table 5  
Avoidance of CG and UA doublets in codons position II-III.

Group	Plants	Dicots	Monocots	Monocots no storage protein	Maize	Soybean	RuBPC SSU	CAB
XCG/XCC	40	30	61	62	67	37	18	22
XTA/XTT	37	35	47	34	43	41	9	13

XCG/XCC and XTA/XAA values are multiplied by 100.

Table 6  
Codon frequencies in 17 plant CAB genes.

	MZW LHC	WIT CAB	ATH LHC	CUS LHC	LGI LHC	LYS <sup>1</sup> LUT	LYS <sup>2</sup> LUT	LYS <sup>3</sup> LUT	LYS <sup>4</sup> LUT	LYS <sup>5</sup> LUT	PET CAB	PET 22L	PET 22R	PET CAB	PET 25	PET CAB	PET 91R	PEA CAB
Gly	5	5	2	2	12	1	3	1	2	2	0	1	2	2	2	0	1	1
Gly	1	0	18	9	0	12	11	8	12	11	12	11	12	12	13	15	10	12
Gly	1	11	8	13	0	13	16	4	13	14	16	13	14	15	15	15	16	13
Gly	25	15	5	7	20	5	2	3	7	4	3	6	2	3	4	4	2	5
Glu	15	13	12	12	15	9	12	6	7	8	11	10	11	10	8	8	11	8
Glu	0	0	3	3	0	7	3	2	10	8	3	5	5	4	8	5	5	7
Asp	0	1	3	3	0	3	2	1	3	4	5	5	5	5	5	4	3	6
Asp	11	11	8	8	13	7	9	4	8	7	6	5	5	6	7	8	8	5
Val	7	7	2	6	5	3	2	1	3	2	3	4	3	4	4	4	3	3
Val	0	0	1	0	0	2	2	1	1	1	3	4	2	2	2	1	2	2
Val	0	4	8	7	1	6	9	7	13	9	9	3	8	9	10	9	9	6
Val	11	6	7	4	11	7	5	4	1	3	6	9	5	5	5	1	6	5
Ala	10	6	1	0	5	0	1	0	0	0	1	0	0	1	0	1	1	0
Ala	0	2	3	2	0	6	1	7	2	4	4	2	4	4	7	6	7	7
Ala	2	5	13	16	0	18	19	11	17	14	17	18	20	14	13	14	14	14
Ala	22	18	16	9	26	9	9	8	6	4	0	10	7	13	9	10	7	7
Arg	0	0	4	2	0	1	1	1	1	2	1	1	0	1	1	4	0	2
Arg	0	0	2	1	1	3	3	2	3	1	3	3	3	3	2	1	3	2
Ser	0	0	1	0	0	1	0	0	0	1	1	2	3	2	2	1	2	1
Ser	8	7	7	6	4	6	6	5	3	3	4	4	5	4	4	3	4	5
Lys	13	11	12	11	12	10	10	7	9	7	11	13	12	12	11	11	14	11
Lys	0	2	0	2	0	2	4	1	4	3	4	2	3	2	3	2	1	3
Asn	0	2	1	6	0	3	5	2	2	3	6	5	4	5	4	3	1	1
Asn	9	8	10	5	9	7	5	6	10	6	5	5	6	6	6	7	9	9
Met	7	7	7	5	6	7	7	6	6	5	7	7	7	7	7	6	7	7
Ile	0	1	1	0	0	0	0	0	0	3	2	0	0	1	0	1	1	1
Ile	1	5	3	3	0	3	5	2	11	6	1	2	5	4	7	1	2	2
Ile	7	4	3	5	12	6	3	4	3	3	5	5	4	4	6	5	4	4
Thr	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
Thr	0	0	5	1	0	2	3	1	3	4	2	1	3	2	4	3	1	1
Thr	0	2	2	4	0	2	4	1	5	4	5	7	5	6	6	4	4	4
Thr	4	6	3	4	4	3	5	1	1	1	4	4	3	2	1	3	3	5

Absolute frequencies of each codon are calculated from sequences in GenBank release 55 or the literature (see Tables 1 and 2). Sequences not in GenBank release 55 are CAB genes referenced in Table 2 as follows:

*Lycopersicon esculentum* CAB 5



Table 7

WHT		MZR	CRE	CRE	CUS	FTR	SOY	INN	LGI	TOM	TOM	TOM	TOM	MES	RAP	TOB	PET	PET	PEA	POT
RBC	B	BCS	RBC	RBC	SSU	BCR	RUB	RUB	3BPC	RBC	SB	RBC	SC	RBC	SAT	RBP	RBC	RBC	RUB	RBC
Gly	3	1	0	0	0	0	0	0	3	1	1	1	1	1	1	0	1	1	1	1
Gly	1	1	0	0	7	7	4	5	0	0	6	0	7	6	8	6	5	4	7	8
Gly	0	0	0	0	1	2	3	4	0	2	2	2	2	4	0	0	3	2	3	2
Gly	3	9	4	4	3	2	7	1	3	0	1	2	0	0	1	4	2	4	2	0
Glu	9	5	6	6	5	9	6	6	11	8	8	8	8	8	6	7	8	7	5	8
Glu	0	0	0	4	2	3	4	0	3	3	3	3	3	2	4	5	2	3	5	3
Asp	0	0	1	1	2	0	4	1	0	3	3	3	3	1	2	3	3	3	4	2
Asp	5	8	7	7	6	4	3	5	6	3	3	3	3	4	6	2	3	3	2	4
Val	5	5	10	10	4	6	4	3	5	3	4	4	4	5	3	6	5	4	5	4
Val	0	0	0	0	1	0	1	0	0	0	0	1	0	2	1	0	0	4	1	0
Val	0	0	0	0	6	3	3	5	1	8	7	5	8	4	5	7	4	5	7	9
Val	4	6	9	9	3	6	2	3	3	3	3	3	3	6	4	1	4	2	3	2
Ala	2	1	2	2	0	0	0	1	1	2	0	0	2	0	0	1	0	0	1	2
Ala	0	0	0	0	2	2	1	0	0	3	5	6	3	3	2	4	4	4	1	2
Ala	1	1	5	3	9	4	7	7	0	8	6	8	6	6	7	5	8	6	6	6
Ala	4	10	22	23	3	8	5	7	6	4	3	1	4	7	4	6	3	7	3	5
Arg	0	1	0	0	4	0	3	1	1	0	1	1	0	0	0	1	1	3	1	0
Arg	0	1	0	0	0	2	1	3	1	2	2	2	2	1	2	2	3	2	4	2
Ser	0	0	0	0	4	2	0	0	0	1	2	2	1	1	2	2	1	1	1	1
Ser	3	6	2	3	3	1	2	2	2	5	4	3	5	2	4	3	3	3	2	4
Iys	8	9	10	10	8	12	9	10	9	7	8	9	7	10	11	10	11	10	10	7
Iys	0	1	0	0	2	2	3	0	3	3	3	1	3	1	2	1	0	1	3	3
Asn	1	0	0	0	2	0	2	1	0	3	4	3	3	3	0	2	4	3	1	3
Asn	2	7	8	8	6	7	5	5	4	7	4	4	7	6	6	7	3	3	3	7
Met	3	6	9	9	6	6	7	7	3	6	6	7	6	7	5	5	5	6	6	6
Ile	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ile	1	0	1	2	4	1	5	1	0	5	6	4	5	1	5	4	4	5	4	5
Ile	5	5	6	5	7	4	5	5	6	3	3	4	3	5	4	4	5	3	3	3
Thr	1	3	0	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Thr	2	0	0	0	0	0	0	1	0	2	4	0	2	1	2	0	0	1	3	1
Ibr	0	0	1	0	4	3	3	3	0	2	5	4	2	2	4	5	7	5	5	4
Thr	1	6	5	4	3	5	6	7	3	5	2	5	5	7	6	3	2	4	2	1
Trp	3	4	4	5	4	5	4	5	4	5	5	5	5	4	4	5	5	5	1	5
Trp	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Absolute frequencies of each codon are calculated from sequences in GenBank release 55 or the literature (see Table 1 and 2). Sequences not in GenBank release 55 are RuBPC SSU genes referenced in Table 2 as follows:**

**! Assembléens bannihem un onvstallinnu RuBPC SSU**

<sup>2</sup> *Raphanus sativus* RUBICSSU

codon position III. However, the soybean codon usage pattern is almost identical to the general dicot pattern, even though it represents a much smaller portion of the entire dicot sample.

In order to determine whether the coding strategy of highly expressed genes such as RuBPC SSU and CAB is more biased than that of plant genes in general, we calculated codon usage profiles for subsets of these genes (20 and 17 sequences respectively) (Table 4). The RuBPC SSU and CAB pooled samples are characterized by stronger avoidance of the codons XCG and XTA than in the larger monocot and dicot samples (Tables 4 and 5). Although most of the genes in these subsamples are dicot in origin (17/20 and 15/17), their codon profile resembles that of the monocots in that G+C is preferred in the degenerate base III.

The use of pooled data for highly expressed genes may obscure identification of species-specific patterns in codon choice. Therefore, we have tabulated the codon choices of individual genes for RuBPC SSU (Table 6) and CAB (Table 7). The preferred codons of the maize and wheat genes for RuBPC SSU and CAB are more restricted in general than are those of the dicot species. Matsuoka et al. (28) noted the extreme codon bias of the maize RuBPC SSU gene as well as two other highly expressed genes in maize leaves, CAB and phosphoenolpyruvate carboxylase. These genes almost completely avoid the use of A+T in codon position III, although this codon bias was not as pronounced in non-leaf proteins such as ADH, zein 22 kDa subunit, sucrose synthetase and ATP/ADP translocator. Since the wheat SSU and CAB genes have a similar pattern of codon preference, this may reflect a common monocot pattern for these highly expressed genes in leaves. The CAB gene for *Lemna* and the RuBPC SSU genes for *Chlamydomonas* share a similar extreme preference for G+C in codon position III.

In dicot CAB genes, however, A+T degenerate bases are preferred by some synonymous codons (e.g. GCT for Ala, CTT for Leu, GGA and GGT for Gly). In general the G+C preference in position III is less pronounced for both RuBPC SSU and CAB genes in dicots than in monocots.

### **Discussion:**

Because of the degenerate nature of the genetic code, only part of the variation contained in a gene is expressed in its protein. It is clear that variation between degenerate base frequencies is not a neutral phenomenon since systematic codon preferences have been reported for bacterial, yeast and mammalian genes. Analysis of a large group of plant gene sequences indicates that synonymous codons are used differently by monocots and dicots. These patterns are also distinct from those reported for *E. coli*, yeast and man (13,18).

In general, the plant codon usage pattern more closely resembles that of man and other higher eukaryotes than unicellular organisms, due to the overall preference for G+C content in codon position III (13,18). Monocots in this sample share the most commonly used codon for 13 of 18 amino acids as that reported for a sample of human genes (18), although dicots favor the most commonly used human codon in only 7 of 18 amino acids.

Several earlier discussions of plant codon usage have focussed on the differences between codon choice in plant nuclear genes and in chloroplasts (18,23). Chloroplasts differ from the nuclear genome of higher plants in that they encode only 30 tRNA species (29, 30). Since

chloroplasts have restricted their tRNA genes, the use of preferred codons by chloroplast-encoded proteins appears more extreme. However, a positive correlation has been reported between the level of isoaccepting tRNA for a given amino acid and the frequency with which this codon is used in the chloroplast genome (31).

Our analysis of the expanded plant sample confirms earlier reports that the nuclear and chloroplast genomes in plants have distinct coding strategies. The codon usage of monocots in this sample is distinct from chloroplast usage, sharing the most commonly used codon for only 1 of 18 amino acids. Dicots in this sample share the most commonly used codon of chloroplasts in only 4 of 18 amino acids. In general, the chloroplast codon profile more closely resembles that of unicellular organisms, with a strong bias towards the use of A+T in the degenerate third base.

In unicellular organisms, highly expressed genes use a smaller subset of codons than do weakly expressed genes, although yeast and *E. coli* use distinct preferred codons in some cases. Sharp and Li (12) report that codon usage in 165 *E. coli* genes reveals a positive correlation between high expression and increased codon bias. Bennetzen and Hall (15) and others (12-14, 17, 18) have described a similar trend in codon selection in yeast. Codon usage in these highly expressed genes correlates with the abundance of isoaccepting tRNAs in both yeast and *E. coli*. If, as Ikemura (16) has proposed, the good fit of abundant yeast and *E. coli* mRNA codon usage to isoacceptor tRNA abundance promotes high translation levels and high steady state levels of these proteins, then the potential for high levels of expression of plant genes in yeast or *E. coli* could be limited by their codon usage. Hoekema et al. (24) report that replacement of the 25 most favored yeast codons with rare codons in the 5' end of the highly expressed gene PGK1 leads to a decrease in both mRNA and protein. These results indicate that codon bias should be considered when engineering high expression of foreign genes in yeast and other systems.

A number of researchers have attempted to express plant genes in yeast (32-34) and *E. coli* (35-37). In the case of wheat  $\alpha$ -gliadin (32),  $\alpha$ -amylase (33) genes, and maize zein genes (34), low levels of expression have been reported in yeast. Neill et al. (32) have suggested that the low levels of expression of  $\alpha$ -gliadin in yeast may be due in part to codon usage bias, since  $\alpha$ -gliadin codons for Phe, Leu, Ser, Gly, Tyr and especially Glu do not correlate well with the abundant yeast isoacceptor tRNAs. In *E. coli*, however, soybean glycinin A2 (35) and wheat RuBPC SSU (36, 37) are expressed adequately.

Not much is known about the makeup of tRNA populations in plants. Viotti et al. (38) report that maize endosperm actively synthesizing zein, a storage protein rich in glutamine, leucine, and alanine, is characterized by higher levels of accepting activity for these three amino acids than are maize embryo tRNAs. This may indicate that the tRNA population of specific plant tissues may be adapted for optimum translation of highly expressed proteins such as zein. To our knowledge, no one has experimentally altered codon bias in highly expressed plant genes to determine possible effects of the protein translation in plants to check the effects on the level of expression. Our data indicate that the highly biased RuBPC SSU and CAB genes would be good candidates for such an experiment.

#### **Acknowledgements:**

We would like to thank Sharon Locken for bibliographic assistance and Lynette Bausch, Lynda Brereton and Sarah Weske for assistance in data entry.

\*To whom correspondence should be addressed

# References

1. Grantham, R. (1980) Trends Biochem. Sci. 5, 327-331.
2. Grantham, R., Gautier, C. and Gouy, M. (1980) Nucl. Acids Res. 8, 1893-1912.
3. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) Nucl. Acids Res. 8, r49-r62.
4. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
5. Aota, S.-I., Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) Nucl. Acids Res. 16, r315-r402.
6. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
7. Gouy, M. and Gautier, C. (1982) Nucl. Acids Res. 10, 7055-7074.
8. Sharp, P.M. and Li, W.-H. (1986) J. Mol. Evol. 24, 28-38.
9. Ogasawara, N. (1985) Gene 40, 145-150.
10. McConnell, D.J., Catwell, B.A., Devine, K.M., Forage, A.J., Laiode, B.M., O'Kane, C., Ollington, J.F. and Sharp, P.M. (1986) Ann. NY Acad. Sci. 469, 1-17.
11. Winans, S.C., Allenza, P., Stachel, S.E., McBride, K.E. and Nester, E.W. (1987) Nucl. Acids. Res. 15, 825-837.
12. Sharp, P.M. and Li, W.-H. (1986) Nucl. Acids Res. 14, 7734-7749.
13. Ikemura, T. (1985) Mol. Biol. Evol. 2, 13-34.
14. Boer, H.A., de and Kastelein, R.A. (1986) In: From Gene to Protein: Steps Dictating the Maximal Level of Gene Expression (eds. J. Davis, B. Reznikoff and L. Gold). Butterworths, New York.
15. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
16. Ikemura, T. (1982) J. Mol. Biol. 158, 573-597.
17. Sharp, P.M., Touhy, T.M.F. and Mosurski, K.R. (1986) Nucl. Acids Res. 14, 5125-5143.
18. Grantham, R., Perrin, P. and Mouchiroud, D. (1986) Oxford Surveys in Evol. Biol. 3, 48-81.
19. Russell, D.W., Smith, M., Williamson, V.M. and Young, E.T. (1983) J. Biol. Chem. 258, 2674-2682.
20. Hastings, K.E.M. and Emerson, C.P. (1983) J. Mol. Evol. 19, 214-218.
21. Bernardi, G. and Bernardi, G. (1985) J. Mol. Evol. 22, 363-365.
22. Lycett, G.W., Delauney, A.J. and Croy, R.R.D. (1983) FEBS Letters 153, 43-46.
23. Boudraa, M. (1987) Genet. Sel. Evol. 19, 143-154.
24. Hoekma, A., Kastelein, R.A., Vasser, M.A. and Boer, H.A. de (1987) Mol. Cell. Biol. 7, 2914-2924.
25. Martinez, H.M., Katzung, B. and Farrah, T. (1984) Sequence Analysis Programs Manual, University of California, San Francisco.
26. Devereux, J. and Haerberli, P. (1983) Program Library of the University of Wisconsin Genetics Computer Group, Madison, Wisconsin.
27. Grantham, R., Greenland, T., Louail, S., Mouchiroud, D., Prato, J.L., Gouy, M. and Gautier, C. (1985) Bull. Inst. Pasteur. 83, 95-148.
28. Matsuka, M., Kano-Murakami, Y., Tanaka, Y., Ozeki, Y. and Yamamoto, N. (1987) J. Biochem. 102, 673-676.
29. Wakasugi, T., Ohme, M., Shinozaki, K. and Sugiura, M. (1986) Plant Mol. Biol. 7, 385-392.
30. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chonwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043-2049.
31. Pfitzinger, H., Guillemaut, P., Weil, J.-H. and Pillay, D.T.N. (1987) Nucl. Acids Res. 15, 1377-1386.

32. Neill, J.D., Litts, J.C., Anderson, O.D., Greene, F.C. and Stiles, J.I. (1987) *Gene* 55, 303-317.
33. Rothstein, S.J., Lahners, K.N., Lazarus, C.M., Baulcombe, D.C. and Gatenby, A.A. (1987) *Gene* 55, 353-356.
34. Coraggio, I., Compagno, C., Martegani, E., Ranzi, B.M., Sala, E., Alberghina, L. and Viotti, A. (1986) *EMBO J.* 5, 459-465.
35. Fuzakawa, C., Udaka, K., Murayama, A., Higuchi, W. and Totsuka, A. (1987) *FEBS Lett.* 224, 125-127.
36. Vies, S.M. van der, Bradley, D. and Gatenby, A.A. (1986) *EMBO J.* 5, 2439-2444.
37. Gatenby, A.A., Vies, S.M. van der, and Rothstein, S.J. (1987) *Eur. J. Biochem.* 168, 227-231.
38. Viotti, A., Balducci, C. and Weil, J.H. (1978) *Biochim. Biophys. Acta* 517, 125-132.

#### Appendix I: References for Tables 1 and 2.

1. Hershey, H.P., Barker, R.F., Idler, K.B., Lissemore, J.L. and Quail, P.H. (1985) *Nucl. Acids Res.* 13, 8543-8559.
2. Rogers, J.C. Dean, D. and Heck, G.R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 6512-6516.
3. Knox, C.A.P., Sonthayanon, B., Chandra, G.R. and Muthukrishnan, S. (1987) *Plant Mol. Biol.* 9, 3-17.
4. Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N. and Miflin, B.J. (1985) *EMBO J.* 4, 9-15.
5. Fincher, G.B., Lock, P.A., Morgan, M.M., Lingelbach, K., Wettenhall, R.E.H., Mercer, J.F.B., Brandt, A. and Thomsen, K.K. (1986) *Proc. Nat. Acad. Sci. U.S.A.* 83, 2081-2085.
6. Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. (1985) *Nucl. Acids Res.* 13, 7327-7339.
7. Mundy, J. and Rogers, J.C. (1986) *Planta* 169, 51-63.
8. Ponz, F., Paz-Ares, J., Hernandez-Lucas, C., Garcia-Olmedo, F. and Carbonero, P. (1986) *Eur. J. Biochem.* 156, 131-135.
9. Gausing, K. and Barkardottir, R. (1986) *Eur. J. Biochem.* 158, 57-62.
10. Chojecki, J. (1986) *Carlsberg Res. Commun.* 51, 211-217.
11. Bohlmann, H. and Apel, K. (1987) *Mol. Gen. Genet.* 207, 446-454.
12. Nielsen, P.S. and Gausing, K. (1987) *FEBS Lett.* 225, 159-162.
13. Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K. (1987) *FEBS Lett.* 221, 43-47.
14. Higuchi, W. and Fukazawa, C. (1987) *Gene* 55, 245-253.
15. Baulcombe, D.C., Huttly, A.K., Martienssen, Barker, R.F. and Jarvis, M.G. (1987) *Mol. Gen. Genet.* 209, 33-40.
16. Lamppa, G.K., Morelli, G. and Chua, N.-H. (1985) *Mol. Cell Biol.* 5, 1370-1378.
17. Litts, J.C., Colwell, G.W., Chakerian, R.L. and Quatrano, R.S. (1987) *Nucl. Acids Res.* 15, 3607-3618.
18. Baulcombe, D.C., Barker, R.F. and Jarvis, M.G. (1987) *J. Biol. Chem.* 262, 13726-13735.
19. Rafalski, J.A. (1986) *Gene* 43, 221-229.
20. Okita, T.W., Cheesbrough, V. and Reeves, C.D. (1985) *J. Biol. Chem.* 260 8203-8213.
21. Sugiyama, T., Rafalski, A., Peterson, D. and Sö ll, D. (1985) *Nucl. Acids Res.* 13, 8729-8737.
22. Tabata, T., Fukasawa, M. and Iwabuchi, M. (1984) *Mol. Gen. Genet.* 196, 397-400.
23. Tabata, T. and Iwabuchi, M. (1984) *Gene* 31, 285-289.
24. Smith, S.M., Bedbrook, J. and Speirs, J. (1983) *Nucl. Acids Res.* 11, 8719-8734.
25. Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. (1985) *J. Mol. Biol.* 183, 499-502.

26. Schwarz-Sommer, Z., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclercq, L., Mattes, M., Berndtgen, R., Peterson, P.A. and Saedler, H. (1987) *EMBO J.* 6, 287-294.
27. Shah, D.M., Hightower, R.C. and Meagher, R.B. (1983) *J. Mol. Appl. Genet.* 2, 111-126.
28. Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J. and Peacock, W.J. (1985) *Nucl. Acids Res.* 13, 727-743.
29. Kelley, P.M. and Tolan, D.R. (1986) *Plant Physiol.* 82, 1076-1080.
30. Baker, A. and Leaver, C.J. (1985) *Nucl. Acids Res.* 13, 5857-5867.
31. Prat, S., Cortadas, J., Puigdomènech, P. and Palau, J. (1985) *Nucl. Acids Res.* 13, 1493-1504.
32. Moore, R.E., Davies, M.S., O'Connell, K.M., Harding, E.I., Wiegand, R.C. and Tiemeier, D.C. (1986) *Nucl. Acids Res.* 14, 7227-7235.
33. Chaubet, N., Philipps, G., Chaboute, M.-E., Ehling, M. and Gigot, C. (1986) *Plant Mol. Biol.* 6, 253-263.
34. Philipps, G., Chaubet, N., Chaboute, M.-E., Ehling, M. and Gigot, C. (1986) *Gene* 42, 225-229.
35. Rochester, D.E., Winer, J.A. and Shah, D.M. (1986) *EMBO J.* 5, 451-458.
36. Matsuoka, M., Kano-Murakami, Y. and Yamamoto, N. (1987) *Nucl. Acids Res.* 15, 6302.
37. Vance, V.B. and Huang, A.H.C. (1987) *J. Biol. Chem.* 262, 11275-11279.
38. Izui, K., Ishijima, S., Yamaguchi, Y., Katagiri, F., Murata, T., Shigesada, K., Sugiyama, T. and Katsuki, H. (1986) *Nucl. Acids Res.* 14, 1615-1628.
39. Lebrun, M., Waksman, G. and Freyssinet, G. (1987) *Nucl. Acids Res.* 15, 4360.
40. Werr, W., Frommer, W.-B., Maas, C. and Starlinger, P. (1985) *EMBO J.* 4, 1373-1380.
41. Marchionni, M. and Gilbert, W. (1986) *Cell* 46, 133-141.
42. Geraghty, D.E., Messing, J. and Rubenstein, I. (1982) *EMBO J.* 1, 1329-1335.
43. Marks, M.D., Lindell, J.S. and Larkins, B.A. (1985) *J. Biol. Chem.* 260, 16451-16459.
44. Prat, S., Pérez-Grau, L. and Puigdomènech, P. (1987) *Gene* 52, 41-49.
45. Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) *Cell* 29, 1015-1026.
46. Marks, M.D. and Larkins, B.A. (1982) *J. Biol. Chem.* 257, 9976-9983.
47. Bethards, L.A., Skadsen, R.W. and Scandalios, J.G. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 6830-6834.
48. Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) *EMBO J.* 6, 3553-3558.
49. Sommer, H. and Saedler, H. (1986) *Mol. Gen. Genet.* 202, 429-434.
50. Chang, C. and Meyerowitz, E.M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 1408-1412.
51. Chaboute, M.-E., Chaubet, N., Philipps, G., Ehling, M. and Gigot, C. (1987) *Plant Mol. Biol.* 8, 179-191.
52. Leutwiler, L.S., Meyerowitz, E.M. and Tobin, E.M. (1986) *Nucl. Acids Res.* 14, 4051-4064.
53. Ludwig, S.R., Oppenheimer, D.G., Silflow, C.D. and Snustad, D.P. (1987) *Proc. Nat. Acad. Sci. U.S.A.* 84, 5833-5837.
54. Klee, H.J., Muskopf, Y.M. and Gasser, C.S. (1987) *Mol. Gen. Genet.* 210, 437-442.
55. Altenbach, S.B., Pearson, K.W., Leung, F.W. and Sun, S.S.M. (1987) *Plant Mol. Biol.* 8, 239-250.
56. Rose, R.E., DeJesus, C.E., Moylan, S.L., Ridge, N.P., Scherer, D.E. and Knauf, V.C. (1987) *Nuc. Acids Res.* 15, 7197.
57. Ericson, M.L., Roedin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. and Rask, L. (1986) *J. Biol. Chem.* 261, 14576-14581.
58. Nasrallah, J.B., Kao, T.-H., Goldberg, M.L. and Nasrallah, M.E. (1985) *Nature* 318, 263-267.

- 
59. Carrington, D.M., Auffret, A. and Hanke, D.E. (1985) *Nature* 313, 64-67.
  60. Cohen, L.W., Coghlan, V.M. and Dihel, L.C. (1986) *Gene* 48, 219-227.
  61. Merchant, S. and Bogorad, L. (1987) *J. Biol. Chem.* 262, 9062-9067.
  62. Goldschmidt-Clermont, M., and Rahire, M. (1986) *J. Mol. Biol.* 191, 421-437.
  63. Sharrock, R.A., Lissemore, J.L. and Quail, P.H. (1986) *Gene* 47, 287-295.
  64. Smith, S.M. and Leaver, C.J. (1986) *Plant Physiol.* 81, 762-767.
  65. Greenland, A.J., Thomas, M.V. and Walden, R.M. (1987) *Planta* 170, 99-110.
  66. Chen, J. and Varner, J.E. (1985) *EMBO J.* 4, 2145-2151.
  67. Chen, J. and Varner, J.E. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4399-4403.
  68. Schnell, D.J. and Etzler, M.E. (1987) *J. Biol. Chem.* 262, 7220-7225.
  69. Adams, C.A., Babcock, M., Leung, F. and Sun, S.M. (1987) *Nucl. Acids Res.* 15, 1875.
  70. Schuler, M.A., Ladin, B.F., Pollaco, J.C., Freyer, G. and Beachy, R.N. (1982) *Nucl. Acids Res.* 10, 8245-8261.
  71. Hammond, R.W., Foard, D.E. and Larkins, B.A. (1984) *J. Biol. Chem.* 259, 9883-9890.
  72. Negoro, T., Momma, T. and Fukazawa, C. (1985) *Nucl. Acids Res.* 13, 6719-6731.
  73. Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K. and Fukazawa, C. (1985) *Eur. J. Biochem.* 149, 491-496.
  74. Fukazawa, C., Momma, T., Hirano, H., Harada, K. and Udaka, K. (1985) *J. Biol. Chem.* 260, 6234-6239.
  75. Momma, T., Negoro, T., Udaka, K. and Fukazawa, C. (1985) *FEBS Lett.* 188, 117-122.
  76. Nagao, R.T., Czarnecka, E., Gurley, W.B., Schöffl, F. and Key, J.L. (1985) *Mol. Cell Biol.* 5, 3417-3428.
  77. Brisson, N. and Verma, D.P.S. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4055-4059.
  78. Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) *Cell* 34, 1023-1031.
  79. Shibata, D., Steczko, J., Dixon, J.E., Hermanson, M., Yazdanparast, R. and Axelrod, B. (1987) *J. Biol. Chem.* 262, 10080-10085.
  80. Sandal, N.N., Bojsen, K. and Marcker, K.A. (1987) *Nucl. Acids Res.* 15, 1507-1519.
  81. Mauro, V.P., Nguyen, T., Katinakis, P. and Verma, D.P.S. (1985) *Nucl. Acids Res.* 13, 239-249.
  82. Katinakis, P. and Verma, D.P.S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4157-4161.
  83. Jacobs, F.A., Zhang, M., Fortin, M.G. and Verma, D.P.S. (1987) *Nucl. Acids Res.* 15, 1271-1280.
  84. Fortin, M.G., Morrison, N.A. and Verma, D.P.S. (1987) *Nucl. Acids Res.* 15, 813-824.
  85. Nguyen, T., Zelechowska, M., Foster, V., Bergmann, H. and Verma, D.P.S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 5040-5044.
  86. Franssen, H.J., Nap, J.-P., Gloudemans, T., Stiekema, W., van Dam, H., Govers, F., Louwerse, J., van Kammen, A., and Bisseling, T. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4495-4499.
  87. Sengupta-Gopalan, C., Pitas, J.W., Thompson, D.V. and Hoffman, L.M. (1986) *Mol. Gen. Genet.* 203, 410-420.
  88. Hong, J.C., Nagao, R.T. and Key, J.L. (1987) *J. Biol. Chem.* 262, 8367-8376.
  89. Berry-Lowe, S.L., McKnight, T.D., Shah, D.M. and Meagher, R.B. (1982) *J. Mol. Appl. Genet.* 1, 483-498.
  90. Krueger, R.W., Holland, M.A., Chisholm, D. and Polacco, J.C. (1987) *Gene* 54, 41-50.
  91. Czarnecka, E., Nagao, R.T., Key, J.L. and Gurley, W.B. (1988) *Mol. Cell Biol.* 8, 1113-1122.
  92. Vierling, E., Nagao, R.T., DeRocher, A.E. and Harris, L.M. (1988) *EMBO J.* 7, 575-581.
  93. Guiltinan, M.J., Ma, D.-P., Barker, R.F., Bustos, M.M., Cyr, R.J., Yadegari, R. and Fosket, D.E. (1988) *Plant Mol. Biol.* 10, 171-184.
  94. Chlan, C.A., Borroto, K., Kamalay, J.A. and Dure, L. (1987) *Plant Mol. Biol.* 9, 533-546.
-



95. Waksman, G. and Freyssinet, G. (1987) *Nucl. Acids Res.* 15, 1328-1328.
96. Allen, R.D., Cohen, E.A., Vonder Haar, R.A., Adams, C.A., Ma, D.P., Nessler, C.L. and Thomas, T.L. (1987) *Mol. Gen. Genet.* 210, 211-218.
97. Sakajo, S., Nakamura, K. and Asahi, T. (1987) *Eur. J. Biochem.* 165, 437-442.
98. Karlin-Neumann, G.A., Kohorn, B.D., Thornber, J.P. and Tobin, E.M. (1985) *J. Mol. Appl. Genet.* 3, 45-61.
99. Stiekema, W.J., Wimpee, C.F. and Tobin, E.M. (1983) *Nucl. Acids Res.* 11, 8051-8061.
100. Konieczny, A. (1987) *Nucl. Acids Res.* 15, 6742.
101. Hoffman, N.E., Pichersky, E., and Cashmore, A.R. (1987) *Nucl. Acids Res.* 15, 3928.
102. Holdsworth, M.J., Bird, C.R., Ray, J., Schuch, W. and Grierson, D. (1987) *Nucl. Acids Res.* 15, 731-739.
103. Grierson, D., Tucker, G.A., Keen, J., Ray, J., Bird, C.R. and Schuch, W. (1986) *Nucl. Acids Res.* 14, 8595-8603.
104. Hoffman, N.E., Pichersky, E., Malik, V.S., Castresana, C., Ko, K., Darr, S.C. and Cashmore, A.R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 8844-8848.
105. Pichersky, E., Bernatzky, R., Tanksley, S.D. and Cashmore, A.R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3880-3884.
106. McKnight, T.D., Alexander, D.C., Babcock, M.S. and Simpson, R.B. (1986) *Gene* 48, 23-32.
107. Lee, J.S., Brown, W.E., Graham, J.S., Pearce, G., Fox, E.A., Dreher, T.W., Ahern, K.G., Pearson, G.D. and Ryan, C.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7277-7281.
108. Graham, J.S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L.H. and Ryan, C.A. (1985) *J. Biol. Chem.* 260, 6561-6564.
109. Pichersky, E., Bernatzky, R., Tanksley, S.D., Breidenbach, R.B., Kausch, A.P. and Cashmore, A.R. (1985) *Gene* 40, 247-258.
110. Pirersky, E., Hoffman, N.E., Malik, V.S., Bernatzky, R., Tanksley, S.D., Szabo, L. and Cashmore, A.R. (1987) *Plant Mol. Biol.* 9, 109-120.
111. Ray, J., Bird, C., Maunders, M., Grierson, D. and Schuch, W. (1987) *Nucl. Acids Res.* 15, 10587.
112. Kiss, G.B., Végh, Z. and Vincze, E. (1987) *Nucl. Acids Res.* 15, 3620.
113. DeRocher, E.J., Ramage, R.T., Michalowski, C.B. and Bohnert, H.J. (1987) *Nuc. Acids Res.* 15, 6301.
114. Boutry, M. and Chua, N.-H. (1985) *EMBO J.* 4, 2159-2165.
115. Calza, R., Huttner, E., Vincentz, M., Rouzé, P., Galangau, F., Vaucheret, H., Chérel, I., Meyer, C., Kronenberger, J., and Caboche, M. (1987) *Mol. Gen. Genet.* 209, 552-562.
116. Tingey, S.V. and Coruzzi, G.M. (1987) *Plant Phys.* 84, 366-373.
117. Shinshi, H., Mohnen, D. and Meins, F. Jr. (1987) *Proc. Natl. Acad. Sci. USA* (1987) 84, 89-93.
118. Shih, M.-C., Lazar, G. and Goodman, H.M. (1986) *Cell* 47, 73-80.
119. Pfitzner, U.M. and Goodman, H.M. (1987) *Nucl. Acids Res.* 15, 4449-4465.
120. Cornelissen, B.J.C., Hooft van Huijsduijnen, R.A.M., Van Loon, L.C. and Bol, J.F. (1986) *EMBO J.* 5, 37-40.
121. Lagrimini, L.M., Burkhart, W., Moyer, M. and Rothstein, S. (1987) *Proc. Natl. Acad. Sci. USA* 84, 7542-7546.
122. Mazur, B.J. and Chui, C.-F. (1985) *Nucl. Acids Res.* 13, 2373-2386.
123. Cornelissen, B.J.C., Hooft van Huijsduijnen, R.A.M. and Bol, J.F. (1986) *Nature* 321, 531-532.
124. Tucker, M.L., Durbin, M.L., Clegg, M.T. and Lewis, L.N. (1987) *Plant Mol. Biol.* 9, 197-203.
125. Reimold, U., Kröger, M., Kreuzaler, F. and Hahlbrock, K. (1983) *EMBO J.* 2, 1801-1805.
126. Dunsmuir, P. (1985) *Nuc. Acids Res.* 13, 2503-2518.

- 
127. Stayton, M.M., Black, M., Bedbrook, J. and Dunsmuir, P. (1986) *Nucl. Acids Res.* 14, 9781-9796.
  128. Koes, R.E., Spelt, C.E., Reif, H.J., van den Elzen, P.J.M., Veltkamp, E. and Mol, J.N.M. (1986) *Nucl. Acids Res.* 14, 5229-5239.
  129. Condit, C.M. and Meagher, R.B. (1986) *Nature* 323, 178-181.
  130. Turner, N.E., Clark, W.G., Tabor, G.J., Hironaka, C.M., Fraley, R.T. and Shah, D.M. (1986) *Nucl. Acids Res.* 14, 3325-3342.
  131. Winter, J., Wright, R., Duck, N., Gasser, C., Fraley, R. and Shah, D. (1988) *Mol. Gen. Genet.* 211, 315-319.
  132. Broglie, K.E., Gaynor, J.J. and Broglie, R.M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 6820-6824.
  133. Hoffman, L.M. and Donaldson, D.D. (1985) *EMBO J.* 4, 883-889.
  134. Gebhardt, C., Oliver, J.E., Forde, B.G., Saarelainen, R. and Mifflin, B.J. (1986) *EMBO J.* 5, 1429-1435.
  135. Lee, J.S. and Verma, D.P.S. (1984) *EMBO J.* 3, 2745-2752.
  136. Hoffman, L.M. (1984) *J. Mol. Appl. Genet.* 2, 447-453.
  137. Edwards, K., Cramer, C.L., Bolwell, G.P., Dixon, R.A., Schuch, W. and Lamb, C.J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 6731-6735.
  138. Slightom, J.L., Drong, R.F., Klassy, R.C. and Hoffman, L.M. (1985) *Nucl. Acids Res.* 13, 6483-6498.
  139. Osborn, T.C., Alexander, D.C., Sun, S.S.M., Cardona, C. and Bliss, F.A. (1988) *Science* 240, 207-210.
  140. Ryder, T.B., Hedrick, S.A., Bell, J.N., Liang, X., Clouse, S.D. and Lamb, C.J. (1987) *Mol. Gen. Genet.* 210, 219-233.
  141. Higgins, T.J.V., Beach, L.R., Spencer, D., Chandler, P.M., Randall, P.J., Blagrove, R.J., Kortt, A.A. and Guthrie, R.E. (1987) *Plant Mol. Biol.* 8, 37-45.
  142. Cashmore, A.R. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2960-2964.
  143. Tingey, S.V., Walker, E.L. and Coruzzi, G.M. (1987) *EMBO J.* 6, 1-9.
  144. Gatehouse, J.A., Brown, D., Evans, I.M., Gatehouse, L.N., Jobes, D., Preston, P. and Croy, R.R.D. (1987) *Nucl. Acids Res.* 15, 7642.
  145. Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) *Nucl. Acids Res.* 12, 4493-4506.
  146. Coruzzi, G., Broglie, R., Edwards, C. and Chua, N.-H. (1984) *EMBO J.* 3, 1671-1679.
  147. Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D. and Boulter, D. (1983) *Nucl. Acids Res.* 11, 2367-2380.
  148. Llewellyn, D.J., Finnegan, E.J., Ellis, J.G., Dennis, E.S. and Peacock, W.J. (1987) *J. Mol. Biol.* 195, 115-123.
  149. Gantt, J.S. and Key, J.L. (1987) *Eur. J. Biochem.* 166, 119-125.
  150. Guidet, F. and Fourcroy, P. (1988) *Nucl. Acids Res.* 16, 2336.
  151. Roberts, L.M., Lamb, F.I., Pappin, D.J.C. and Lord, J.M. (1985) *J. Biol. Chem.* 260, 15682-15686.
  152. Halling, K.C., Halling, A.C., Murray, E.E., Ladin, B.F., Houston, L.L. and Weaver, R.F. (1985) *Nucl. Acids Res.* 13, 8019-8033.
  153. Beeching, J.R. and Northcote, D.H. (1987) *Plant Mol. Biol.* 8, 471-475.
  154. Smeekeens, S., van Binsbergen, J. and Weisbeek, P. (1985) *Nucl. Acids Res.* 13, 3179-3194.
  155. Smeekeens, S., de Groot, M., van Binsbergen, J. and Weisbeek, P. (1985) *Nature* 317, 456-458.
  156. Martin, W. and Creff, R. (1986) *Eur. J. Biochem.* 159, 323-331.
  157. Mignery, G.A., Pikaard, C.S., Hannapel, D.J. and Park, W.D. (1984) *Nucl. Acids Res.* 12, 7987-8000.
  158. Cleveland, T.E., Thornburg, R.W. and Ryan, C.A. (1987) *Plant Mol. Biol.* 8, 199-207.
-

159. Eckes, P., Rosahl, S., Schell, J. and Willmitzer, L (1986) *Mol. Gen. Genet.* 205, 14-22.
160. Keil, M., Sanchez-Serrano, J., Schell, J. and Willmitzer, L. (1986) *Nucl. Acids Res.* 14, 5641-5650.
161. Wolter, F.P., Fritz, C.C., Willmitzer, L., Schell, J. and Schreier, P.H. (1988) *Proc. Natl. Acad. Sci. USA* 85,846-850.
162. Salanoubat, M. and Belliard, G. (1987) *Gene* 60, 47-56.
163. Scherer, D.E. and Knauf, V.C. (1987) *Plant Mol. Biol.* 9, 127-134.
164. Jansen, T., Rother, C., Steppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. and Herrmann, R.G. (1987) *FEBS Lett.* 216, 234-240.
165. Rother, C., Jansen, T., Tyagi, A., Tittgen, J. and Herrmann, R.G. (1986) *Curr. Gen.* 11, 171-176.
166. Tyagi, A., Hermans, J., Steppuhn, J., Jansson, Ch., Vater F. and Herrmann, R.G. (1987) *Mol. Gen. Genet.* 207, 288-293.
167. Volokita, M. and Somerville, C.R. (1987) *J. Biol. Chem.* 262, 15825-15828.
168. Kuhse, J. and Pühler, A. (1987) *Plant Sci.* 49, 137-143.
169. Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C. (1986) *Nucl. Acids Res.* 14, 2707-2720.
170. Bassünér, R., Hai, N.V., Jung, R., Saalbach, G. and Müntz, K. (1987) *Nucl. Acids Res.* 15, 9609.

## Steroid Recognition by Chloramphenicol Acetyltransferase: Engineering and Structural Analysis of a High Affinity Fusidic Acid Binding Site

Iain A. Murray<sup>1</sup>, Pauline A. Cann<sup>1</sup>, Philip J. Day<sup>1</sup>, Jeremy P. Derrick<sup>1</sup>  
Michael J. Sutcliffe<sup>1</sup>, William V. Shaw<sup>1\*</sup> and Andrew G. W. Leslie<sup>2</sup>

<sup>1</sup>Department of Biochemistry  
University of Leicester  
Leicester, LE1 7RH, UK

<sup>2</sup>MRC Laboratory of  
Molecular Biology  
Cambridge, CB2 2QH, UK

The antibiotic fusidic acid and certain closely related steroidal compounds are potent competitive inhibitors of the type I variant of chloramphenicol acetyltransferase (CAT<sub>I</sub>). In the absence of crystallographic data for CAT<sub>I</sub>, the structural determinants of steroid binding were identified by (1) construction *in vitro* of genes encoding chimaeric enzymes containing segments of CAT<sub>I</sub> and the related type III variant (CAT<sub>III</sub>) and (2) site-directed mutagenesis of the gene encoding CAT<sub>III</sub>, followed by kinetic characterisation of the substituted variants. Replacement of four residues of CAT<sub>III</sub> (Gln92, Asn146, Tyr168 and Ile172) by their equivalents from CAT<sub>I</sub> yields an enzyme variant that is susceptible to competitive inhibition by fusidate with respect to chloramphenicol ( $K_i = 5.4 \mu\text{M}$ ). The structure of the complex of fusidate and the Q92C/N146F/Y168F/I172V variant, determined at 2.2 Å resolution by X-ray crystallography, reveals the inhibitor bound deep within the chloramphenicol binding site and in close proximity to the side-chain of His195, an essential catalytic residue. The aromatic side-chain of Phe146 provides a critical hydrophobic surface which interacts with non-polar substituents of the steroid. The remaining three substitutions act in concert both to maintain the appropriate orientation of Phe 146 and *via* additional interactions with the bound inhibitor. The substitution of Gln92 by Cys eliminates a critical hydrogen bond interaction which constrains a surface loop (residues 137 to 142) of wild-type CAT<sub>III</sub> which must move in order for fusidate to bind to the enzyme. Only two hydrogen bonds are observed in the CAT-fusidate complex, involving the 3- $\alpha$ -hydroxyl of the A-ring and both hydroxyl of Tyr25 and NE2 of His195, both of which are also involved in hydrogen bonds with substrate in the CAT<sub>III</sub>-chloramphenicol complex. In the acetyl transfer reaction catalysed by CAT, NE2, of His195 serves as a general base in the abstraction of a proton from the 3-hydroxyl of chloramphenicol as the first chemical step in catalysis. The structure of the CAT-inhibitor complex suggests that deprotonation of the 3- $\alpha$ -hydroxyl of bound fusidate by this mechanism could produce an oxyanion nucleophile analogous to that seen with chloramphenicol, but one which is incorrectly positioned to attack the thioester carbonyl of acetyl-CoA, accounting for the observed failure of CAT to acetylate fusidate.

© 1995 Academic Press Limited

\*Corresponding author

**Keywords:** chloramphenicol acetyltransferase; steroid binding; competitive inhibition; chimaeric enzymes; protein engineering

Present addresses: I. A. Murray, Department of Pharmaceutical Chemistry, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0446, U.S.A.; P. J. Day, Department of Chemistry and Biochemistry, University of Texas at Austin, Welch 5252, Austin, TX 78712, U.S.A.; J. P. Derrick, Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK; M. J. Sutcliffe, Department of Chemistry, University of Leicester, Leicester, LE1 7RH, UK  
Abbreviations used: CAT, chloramphenicol acetyltransferase; Cm, chloramphenicol; FA, fusidic acid.

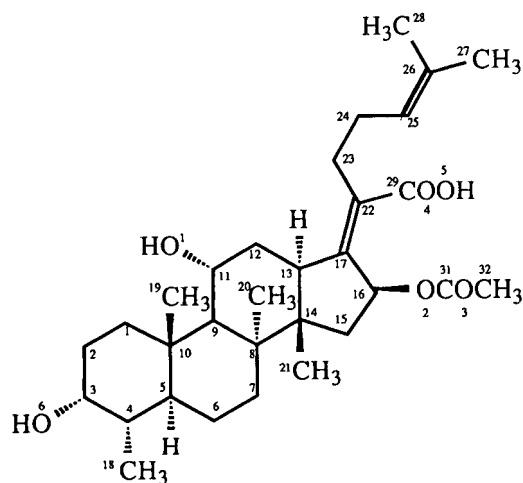
## Introduction

Steroid hormones and related steroidal molecules are known to bind to a range of macromolecular targets in eukaryotes including nuclear receptors, intracellular enzymes involved in steroid biosynthesis or metabolism, and sex-hormone and corticosteroid binding globulins present in plasma (Parker, 1993). Although the structure of the complex between progesterone and an anti-steroid Fab' fragment (Arevalo *et al.*, 1993), and unliganded structures of ketosteroid isomerase (Westbrook *et al.*, 1984), uteroglobin (Morize *et al.*, 1987), and a bacterial  $\beta$ -hydroxysteroid dehydrogenase (Ghosh *et al.*, 1991), have been determined by X-ray crystallography, the breadth of the spectrum of structural motifs for steroid binding by proteins remains unknown. In fact, the sole example of the structural determination of the complex of a steroidal ligand and an *in vivo* protein target is that recently determined for dehydroisoandrosterone bound to a bacterial cholesterol oxidase (Li *et al.*, 1993), a flavin-dependent enzyme believed to have a role in the degradation of 3- $\beta$ -hydroxysteroids, permitting the bacterium to utilise cholesterol as a sole source of carbon and energy.

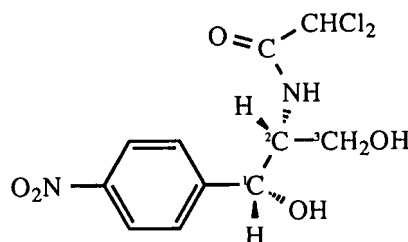
It is well documented that the type I variant (CAT<sub>I</sub>) of the enzyme chloramphenicol acetyltransferase (CAT; EC 2.3.1.28), a common effector of bacterial resistance to the antibiotic chloramphenicol (Shaw, 1967), can also confer resistance to the steroidal antibiotic fusidic acid when expressed in an *Escherichia coli* mutant which, unusually for Gram-negative bacteria, is fusidate-sensitive (Bennett & Shaw, 1983). Resistance to chloramphenicol is achieved *via* the enzyme-catalysed transfer of an acetyl group from acetyl-CoA to the primary hydroxyl group of the substrate, yielding 3-acetylchloramphenicol which fails to bind to, and thus no longer inhibits, the peptidyl-transferase

centre of prokaryotic ribosomes. Fusidic acid is also an inhibitor of bacterial protein synthesis but is known to act at the elongation phase *via* stabilisation of the complex of elongation factor G and GDP (reviewed by Cundliffe, 1981). As binding by CAT<sub>I</sub> leads neither to acetylation nor to any other covalent modification of fusidate, it is likely that the resistance mechanism is simply one of sequestering the antibiotic from its target (Bennett & Shaw, 1983).

Genes encoding CAT have been isolated from many bacterial genera and nucleotide sequences of some 30 different natural variants are known, demonstrating 25 to 98% amino acid identity in pairwise comparisons. None are homologous in primary structure to known steroid-binding proteins, nor, with the exception of the acetyltransferase component of the pyruvate dehydrogenase complex which has the same tertiary fold as CAT (Mattevi *et al.*, 1992), are they known to be related structurally to other proteins of any type. All CAT variants are homotrimers (3  $\times$  25 kDa) with three active sites located in the clefts at the subunit interfaces (Shaw & Leslie, 1991) and operate *via* a rapid-equilibrium ternary complex mechanism with random order of addition of substrates (Kleanthous & Shaw, 1984). The imidazole of His195 acts as a general base to abstract a proton from the primary hydroxyl of chloramphenicol, followed by attack of the resulting oxygen nucleophile at the carbonyl of the acetyl-CoA thioester. The resulting tetrahedral oxyanion intermediate, stabilised by hydrogen-bonding with the hydroxyl of Ser148, then collapses to yield the products 3-acetylchloramphenicol and CoA. Such a mechanism is supported by the results of chemical modification experiments (Kleanthous *et al.*, 1985) as well as X-ray crystallographic (Leslie, 1990) and site-directed mutagenesis (Lewendon *et al.*, 1990, 1994) studies of the type III variant of CAT (CAT<sub>III</sub>) (Murray *et al.*, 1988). In contrast, fusidate resistance appears to be a unique property

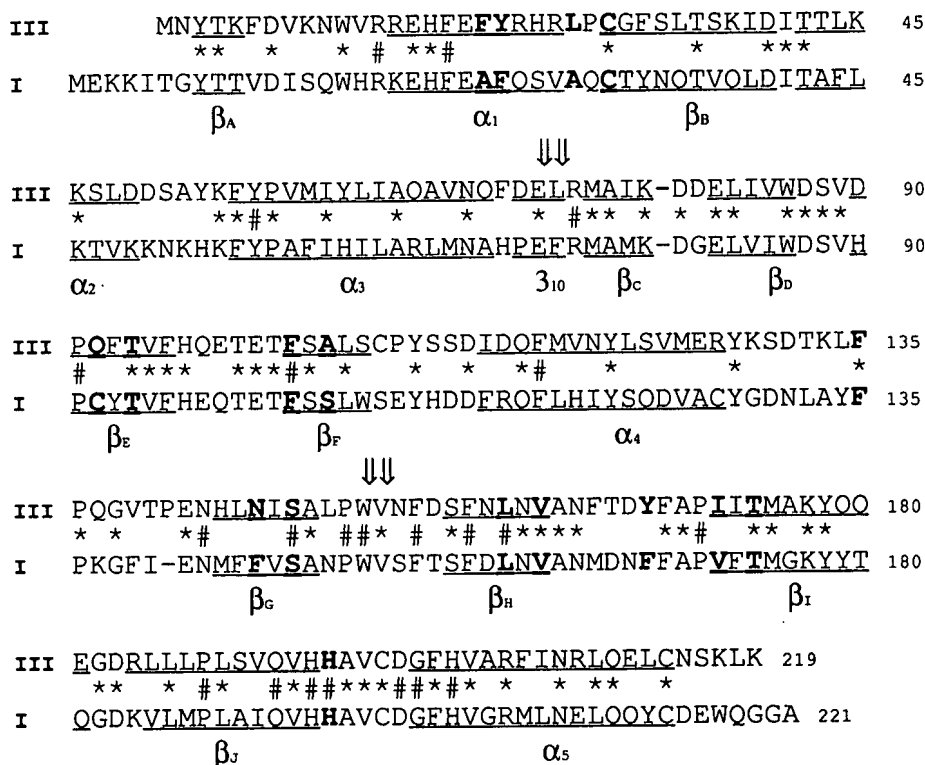


FUSIDIC ACID



CHLORAMPHENICOL

Figure 1. Chemical structure of fusidic acid and chloramphenicol.



**Figure 2.** Sequence alignment of CAT<sub>III</sub> and CAT<sub>I</sub>. The primary sequences of CAT<sub>III</sub> and CAT<sub>I</sub> were optimally aligned and numbered as described by Shaw & Leslie (1991). Secondary structural motifs (helices and strands) observed in the crystal structure of CAT<sub>III</sub> (Leslie, 1990), and the homologous regions of CAT<sub>I</sub>, are underlined. The 17 residues which form the chloramphenicol binding site are shown in bold text. Positions of residue identity are indicated by \* and those amino acids which are conserved in all known CAT variants by #. Double arrows above the sequence indicate the crossover points exploited in the construction of chimaeric CAT variants.

of the CAT<sub>I</sub> enzyme, encoded by the transposon Tn9 and many different enterobacterial R-plasmids. Fusidate binding to CAT<sub>I</sub> is competitive with respect to chloramphenicol, despite the lack of obvious structural equivalence between the two ligands (Figure 1), and extends to various fusidate derivatives and related steroidal molecules (Bennett & Shaw, 1983).

An understanding of the determinants of binding of ligands which are not substrate analogues to proteins is a prerequisite to rational drug design and to the production of novel enzymes by protein engineering. In the case of steroid binding by CAT<sub>I</sub>, such an objective could in principle be achieved by determination of the structure of the enzyme/fusidate complex by X-ray crystallography. However, a prolonged and exhaustive effort has failed to yield crystals of CAT<sub>I</sub> with suitable diffraction properties. Efforts to model the binding of fusidate to CAT<sub>I</sub> using the known structure of the complex of CAT<sub>III</sub> and chloramphenicol (Leslie, 1990; M.J.S., unpublished experiments) have been frustrated by two obstacles; a lack of residue conservation between CAT<sub>I</sub> and CAT<sub>III</sub> (46% identity), even in the substrate binding sites, and the fact that a chloramphenicol site is located deep within each of the three inter-subunit clefts and could, in principle, be occluded (rather than occupied) by fusidate

bound on the trimer surface. Random mutagenesis of the genes encoding either CAT<sub>I</sub> or CAT<sub>III</sub> was considered inappropriate due to the absence of a direct or facile selection for fusidate-sensitive mutants of the former, and the consideration that single point mutations were unlikely to confer fusidate resistance upon *E. coli* expressing CAT<sub>III</sub>. However, starting with the assumption that the tertiary folds of CAT<sub>I</sub> and CAT<sub>III</sub> were likely to be essentially the same, and hence that steroid binding affinity is conferred primarily as a consequence of differences in amino acid side-chains, it seemed reasonable that site-directed mutagenesis of CAT<sub>III</sub> could be useful in mapping the fusidate binding site. The desired result was achieved by the construction *in vitro* of recombinants expressing chimaeric CAT enzymes, followed by directed mutagenesis to produce variants of CAT<sub>III</sub> which bind fusidic acid with high affinity and are amenable to crystallographic analysis.

## Results and Discussion

### Production and characterisation of CAT<sub>I</sub>/CAT<sub>III</sub> chimaeric enzymes

Steady state kinetic analysis of the inhibition of wild-type CAT<sub>I</sub> and CAT<sub>III</sub> by sodium fusidate

indicated that the steroidal inhibitor is bound ~200-fold more strongly by the former ( $K_i = 1.5 \mu\text{M}$  and  $279 \mu\text{M}$ , respectively). In each case binding is competitive with respect to the substrate chloramphenicol. However, because chloramphenicol binds in each of three deep clefts on the enzyme surface of CAT<sub>III</sub> (and presumably CAT<sub>I</sub> also) it cannot be presumed *a priori* that fusidate binding is determined solely by the amino acid side-chains which form the substrate binding pocket. Calculations (using a sphere of radius 1.4 Å) indicated that in the CAT<sub>III</sub>-chloramphenicol complex (Leslie, 1990) only one chlorine atom and the two oxygen atoms of the *p*-nitro substituent are solvent-accessible. Thus, binding of fusidate to non-conserved surface residues flanking the entrance to the chloramphenicol site (residues 26 to 30, 137 to 142 and 163 to 167; Figure 2) might in principle result in competitive inhibition by simply precluding access of substrate to the catalytic centre of CAT<sub>I</sub>. Of the 17 amino acid side-chains which form the chloramphenicol binding site of CAT<sub>III</sub>, only nine are retained in CAT<sub>I</sub> of which four are conserved in all CAT variants (Figure 2; Zhao & Aoki, 1992). As a first step in delineating regions of CAT<sub>I</sub> which contained determinants of high affinity steroid binding, we constructed a series of recombinant CAT genes encoding chimaeric enzymes.

Two crossover points were selected for construction of CAT<sub>I</sub>/CAT<sub>III</sub> chimaeras at locations which divide the CAT open reading frames approximately into thirds. The first crossover point, in a short  $3_{10}$  helix on the enzyme surface between residues 72 and 73, was chosen because this site is known to tolerate insertions of additional amino acid residues while retaining activity (Betz & Sadler, 1981; I.A.M., unpublished experiments). The second was located between two highly conserved surface residues (Pro151 and Trp152) which form part of a reverse turn between  $\beta$ -strands G and H. The fact that the two CAT variants are 46% identical in sequence and readily form heterotrimers of CAT<sub>I</sub> and CAT<sub>III</sub> subunits that are catalytically competent (Packman & Shaw, 1981; Day *et al.*, 1995) raised the possibility that some of the chimaeric enzymes might be active, notwithstanding the potential for folding defects and/or inappropriate intra- and inter-monomer interactions. Of six possible chimaeric proteins encoded by crossovers at these positions, only five produced CAT protein, four of which were able to acetylate chloramphenicol and three were characterised in respect of inhibition by sodium fusidate. Enzymes in which residues 6 to 71 of CAT<sub>III</sub> replaced residues 1 to 71 of CAT<sub>I</sub> and in which residues 152 to 221 of CAT<sub>I</sub> replaced residues 152 to 219 of CAT<sub>III</sub> (the III-I-I and III-III-I chimaeras) were purified by standard affinity chromatography methods and their steady state kinetic parameters for acetyl transfer to chloramphenicol and inhibition by fusidate were determined (Table 1). The former is inhibited by fusidate ( $K_i = 2.0 \mu\text{M}$ ) almost as effectively as is wild-type CAT<sub>I</sub> ( $1.5 \mu\text{M}$ ), whereas the latter is ~twofold less sensitive to fusidate

**Table 1.** Steady state kinetic parameters determined for wild-type and chimaeric CAT variants

Variant	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_m$ Cm ( $\mu\text{M}$ )	$K_i$ FA ( $\mu\text{M}$ )
Wild-type CAT <sub>I</sub>	97 <sup>a</sup>	11 <sup>a</sup>	1.5
Wild-type CAT <sub>III</sub>	599 <sup>b</sup>	11.6 <sup>b</sup>	279
III-I-I	9.3	19	2.0
III-III-I	191	65	505
III (I <sub>139-143</sub> )	534	17	193

III-I-I, Chimaeric CAT variant wherein residues 1 to 71 of CAT<sub>I</sub> were replaced by residues 6 to 71 of CAT<sub>III</sub>.

III-III-I, Chimaeric CAT variant wherein residues 152 to 219 of CAT<sub>III</sub> were replaced by residues 152 to 221 of CAT<sub>I</sub>.

III (I<sub>139-143</sub>), Chimaeric CAT variant wherein loop 139 to 143 (VTPE) of CAT<sub>III</sub> is replaced by the equivalent loop (FIEN) of CAT<sub>I</sub>.

Kinetic parameters were determined as described in Materials and Methods. Parameters are the mean of a minimum of three independent determinations, and standard error values are 15% (or less) of the quoted values.  $K_m$  values for the binding of the second substrate, acetyl-CoA, are within threefold of the values of the wild-type enzymes in the cases of all CAT variants analysed in the present study (data not shown).

<sup>a</sup> Data from Murray *et al.* (1991b).

<sup>b</sup> Data from Lewendon *et al.* (1988).

( $K_i = 505 \mu\text{M}$ ) than wild-type CAT<sub>III</sub>. A third chimaera (III-I-III) wherein residues 72 to 151 of CAT<sub>III</sub> are replaced by their equivalents from CAT<sub>I</sub> could only be assayed in crude cell-free extracts by means of a very sensitive radiometric assay (Gorman *et al.*, 1982) but did display fusidate sensitivity intermediate to that of the two wild-type enzymes assayed under similar conditions (data not shown). Taken together, such data suggest that the principal determinants of high affinity fusidate binding in CAT<sub>I</sub> cannot reside in the N-terminal third of the protein and, additionally, that the C-terminal third of CAT<sub>I</sub> is insufficient by itself to confer CAT<sub>I</sub>-like affinity for fusidate, only binding the ligand in the context of residues 72 to 151 of that variant. It seemed likely therefore that differences in steroid affinity are most likely a consequence of substitutions of some of those residues (92, 105, 146, 168 and 172) known to be variable in the chloramphenicol binding site and/or in the surface loops 137 to 142 and 163 to 167. However, replacement of residues 139 to 143 (Val-Thr-Pro-Glu-Asn) of CAT<sub>III</sub> by the shorter loop (Phe-Ile-Glu-Asn) from CAT<sub>I</sub> yields an enzyme with kinetic parameters that are almost identical to those of wild-type CAT<sub>III</sub> (Table 1). Therefore, site-directed mutagenesis of each of the eight non-conserved residues in the chloramphenicol binding site of CAT<sub>III</sub> was used in an attempt to further define fusidate binding residues.

#### Site-directed mutagenesis of residues of the chloramphenicol binding site of CAT<sub>III</sub>

Eight of the 17 amino acid side-chains which form the chloramphenicol binding site of CAT<sub>III</sub> are substituted (Xaa) in CAT<sub>I</sub>: Phe24(Ala), Tyr25(Phe), Leu29(Ala), Gln92(Cys), Ala105(Ser), Asn146(Phe), Tyr168(Phe) and Ile172(Val). The Y25F mutant was

**Table 2.** Steady state kinetic parameters determined for singly and multiply substituted CAT<sub>III</sub> variants

Variant	$k_{cat}$ (s <sup>-1</sup> )	$K_m$ Cm ( $\mu$ M)	$K_i$ FA ( $\mu$ M)
A.			
Wild-type CAT <sub>III</sub>	599 <sup>a</sup>	11.6 <sup>a</sup>	279
F24A/L29A	500	33	133
Y25F	258 <sup>b</sup>	15 <sup>b</sup>	111
N146F	690	42	>2500
Y168F/I172V	358	38	255
B.			
Q92C/N146F	243	20	44
Q92C/N146F Y168F/I172V	377	20	5.4
F24A/L29A Q92C/N146F			
Y168F/I172V	112	48	3.8
Wild-type CAT <sub>I</sub>	97 <sup>c</sup>	11 <sup>c</sup>	1.5

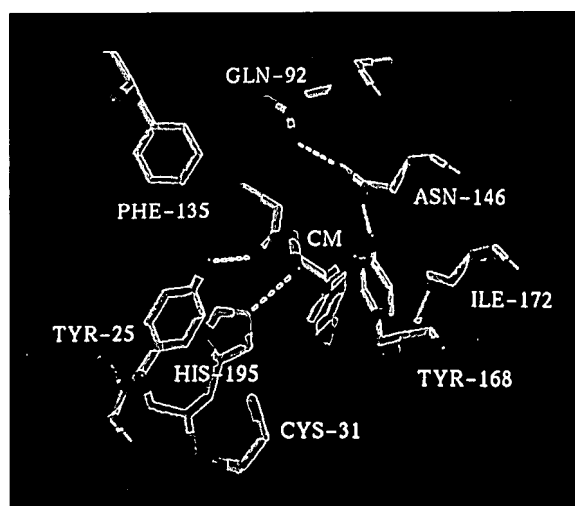
<sup>a</sup> Data from Lewendon *et al.* (1988).<sup>b</sup> Data from Murray *et al.* (1991a).<sup>c</sup> Data from Murray *et al.* (1991b).

prepared and analysed by steady state kinetic methods in an earlier study (Murray *et al.*, 1991a). The substitutions Q92C, A105S and N146F were each introduced into CAT<sub>III</sub> as individual point mutations, whereas the F24A/L29A and Y168F/I172V double mutants were prepared using a single round of mutagenesis taking advantage of the close proximity of the codons of both pairs of residues in the CAT structural gene. The steady state kinetic parameters for chloramphenicol acetylation and inhibition by sodium fusidate for four such variants are presented in Table 2A. Data were not determined for the Cys92 and Ser105 enzymes because preliminary experiments were indicative of a fusidate sensitivity very similar to that of wild-type CAT<sub>III</sub> in each case. With the exception of the N146F variant, each of the residue substitutions has only a minor influence on steroid binding. However, the substitution of Asn146 by phenylalanine effectively abolishes competitive binding of fusidate by CAT<sub>III</sub>. Indeed, whereas inhibition of the N146F enzyme is seen at very high concentrations of inhibitor (>2.5 mM) it is impossible to distinguish between the effect of inhibitor binding to CAT and probable indirect effects of the detergent-like fusidate molecule on the partitioning of chloramphenicol between enzyme and solvent water (data not shown). Loss of affinity for the inhibitor accompanying the N146F mutation does not appear to be a consequence of any global changes in the integrity of the enzyme as judged by the near-wild-type kinetic properties of N146F in the acetyl transfer reaction. Therefore, and in spite of the fact that the N146F substitution has an effect opposite to that predicted, it seemed plausible that Phe146 might have an important role in binding of the inhibitor. In the structure of the binary complex of CAT<sub>III</sub> and chloramphenicol, the side-chain of Asn146 is the centre of an extended hydrogen bond network which includes the amide side-chain of Gln92 and the phenolic hydroxyl of Tyr168 (Fig-

ure 3). As each of these hydrogen bond acceptors and donors is substituted by more hydrophobic amino acids (Cys92, Phe146, Phe168) in CAT<sub>I</sub> it seemed likely that concerted substitution of all three residues might be necessary to produce a CAT<sub>I</sub>-like fusidate binding phenotype. To that end we constructed and purified the Q92C/N146F double and Q92C/N146F/Y168F/I172V quadruple mutants of CAT<sub>III</sub> and determined their kinetic parameters (Table 2B). Introduction of Cys92 in the context of Phe146 enhances binding >50-fold compared to N146F, yielding an enzyme with sixfold greater fusidate ( $K_i = 44 \mu$ M) affinity than that of wild-type CAT<sub>III</sub>. When these two substitutions are combined with those of Y168F and I172V the fusidate inhibition constant ( $K_i$ ) falls to 5.4  $\mu$ M, a 50-fold enhancement in affinity over that of wild-type CAT<sub>III</sub> and approaching that of the type I variant. Further modification via the F24A and L29A substitutions results in only a modest increase in binding affinity ( $K_i = 3.8 \mu$ M) supporting the conclusion of the chimaeric enzyme experiments that residues of the N-terminal segment of CAT<sub>I</sub> play a limited role in steroid binding.

#### Crystal structure of the fusidate-CAT<sub>III</sub>(Q92C/N146F/Y168F/I172V) complex

The structure of the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub> was determined at 2.2 Å resolution in the presence of bound fusidate (Table 3). With the exception of the substituted residues and the different bound ligands the structure is essentially



**Figure 3.** Hydrogen bond interactions in the complex of wild-type CAT<sub>III</sub> and chloramphenicol. NE2 of His195, a critical catalytic residue, acts as a general base to abstract the hydroxyl proton of chloramphenicol as the initial chemical step of the acetyl-transfer reaction. Residues Gln92, Asn146 and Tyr168, which form an extended hydrogen bond network on one side of the substrate binding site (Leslie, 1990), are each replaced by hydrophobic amino acid side-chains in CAT<sub>I</sub>.



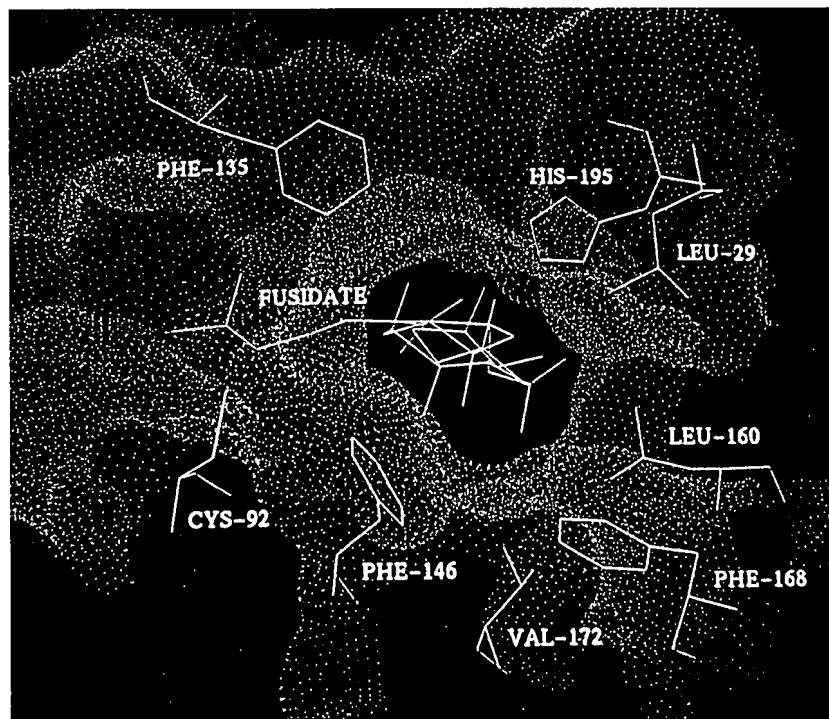


Figure 4. Fusidate bound in the chloramphenicol binding site of the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub>. The solvent-accessible surface (sphere radius 1.5 Å) was calculated for all amino acid side-chains which contained atoms positioned within 6 Å of atoms of the fusidate ligand. The four rings of the steroidal ligand are bound in a roughly cylindrical hydrophobic cavity with the A-ring at the base and D-ring closest to the surface of the trimer. The hydrophobic tail of the inhibitor projects out of the binding site onto the enzyme surface.

identical to that of wild-type CAT<sub>III</sub> in complex with chloramphenicol (Leslie, 1990). The RMS differences between the structures of wild-type and quadruply substituted variants are 0.20 Å and 0.53 Å for main-chain and side-chain atoms, respectively. Two residues, Val139 and Thr140, are disordered in the structure and were deleted from the model. Such disorder reflects the loss of a stabilising hydrogen bond interaction between the side chains of Gln92 and Thr140 (which accompanies the Q92C substitution) and is also observed in the complex between the quadruple mutant and chloramphenicol (A.G.W.L., unpublished data). In addition the side-chain of His144 occupies two distinct positions due to alternate  $\chi_1$  angles of +60° (c.f. wild-type CAT<sub>III</sub>) and -70°. As the peptide bond of His144 shows the highest deviation from planarity (12°) of any residue in wild-type CAT<sub>III</sub>, the alternate conformer of this side-chain may reflect the loss of some conformational constraint following substitution of Asn146 by phenylalanine in the mutant.

The side-chains of Cys92, Phe146, Phe168 and Val172 and the atoms of the four ring systems of the fusidate molecule are well ordered in the CAT-inhibitor complex. The steroid A-ring is positioned at the base of the chloramphenicol binding pocket in close proximity to the imidazole side-chain of His195, an essential catalytic residue (Lewendon *et al.*, 1994). The B, C and D rings also occupy the

chloramphenicol site whereas the remainder of the inhibitor projects onto the enzyme surface and is more accessible to solvent. Atoms of the hydrophobic "tail", the O-acetyl, and carboxylate substituents of the inhibitor (Figure 1) are disordered. In spite of the apparent mobility of the latter it may contribute to fusidate binding via electrostatic interaction with the guanidinium of Arg28, a side-chain which occupies clearly defined density in this structure but is disordered beyond CB in crystals of wild-type CAT<sub>III</sub>.

As might be expected for a hydrophobic steroidal ligand, extensive non-polar interactions are observed in the CAT-fusidate complex (Figure 4). The C-14 methyl substituent (C-21) of the C-ring is in van der Waals contact (<3.6 Å) with all six carbon atoms of the side-chain of Phe146. CD2 and CE2 of the same residue also interact with the C-12 methylene group of the C-ring, supporting the earlier supposition that Phe146 is an important determinant of fusidate binding. CD2 of Leu160 contacts the C-4 methyl substituent (C-18) on the A-ring of fusidate and CD2 of Leu29 is involved in a similar interaction with the C-15 methylene group of the D-ring. In all, there are 57 contact distances of <4 Å between atoms of fusidate and the protein (Table 4) and a further nine to four ordered water molecules in the binding site. Perhaps surprisingly, none of the remaining three substituted residues (Cys92, Phe168 and Val172) make direct contact

Table 3. Refinement statistics for the CAT-fusidate complex

Diffraction data		
Reflections (6.0 to 2.2 Å)	13,517 (100% complete)	
R-value	17.4%	
Atomic model	No. of atoms	Mean isotropic thermal parameter (Å <sup>2</sup> )
Protein	1686	19.7
Fusidate	37	41.0
Solvent <sup>a</sup>	146	37.5
Stereochemical refinement parameter	RMS deviation from ideal values	Refinement restraint weighting values
Bond distances (Å)	0.018	0.020
Bond angles (Å)	0.049	0.040
Planar 1-4 distances (Å)	0.059	0.050
Planes (Å)	0.015	0.020
Chiral volumes (Å <sup>3</sup> )	0.166	0.150

<sup>a</sup> Includes 144 water atoms and two cobalt ions.

with the bound inhibitor. However the side-chains of Cys92 and Phe168 almost certainly contribute to the van der Waals binding energy via interactions with C-26/C-27 of the hydrophobic tail and atoms of the C-16 (O-acetyl) substituent, which are poorly ordered in the crystal. It is also probable that these substitutions are required to permit the appropriate orientation of Phe146 and to provide a more hydrophobic environment for binding of the apolar ring systems. The observed abolition of fusidate binding when the N146F substitution is made in isolation most likely reflects the inappropriateness of introducing a hydrophobic side-chain in the context of the relatively hydrophilic environment provided by Gln92 and Tyr168 of CAT<sub>III</sub>. In wild-type CAT<sub>III</sub> the amide nitrogen of Gln92 is involved in a hydrogen bond interaction with the side-chain hydroxyl of Thr140 stabilising the conformation of a surface loop (residues 137 to 142) adjacent to the chloramphenicol binding site. Elimination of this H-bond, via the Q92C substitution, facilitates a major re-orientation of the loop which would otherwise preclude binding of the hydrophobic tail of fusidate (Figure 5). The truncation of Tyr168 to phenylalanine is required to avoid a steric clash with the new position of residue 146 and is accompanied by movement of the entire side-chain of Phe168 towards residue 172. This in turn necessitates the substitution of Ile172 by the shorter valine side-chain.

The 3- $\alpha$ -hydroxyl group (O-6) of the A-ring is involved in hydrogen bonds with both NE2 of His195 (2.4 Å) and the phenolic-OH of Tyr25 (2.8 Å; Figure 6). Both Tyr25 and His195 are also involved in hydrogen bond interactions with chloramphenicol when it is bound to wild-type CAT<sub>III</sub> (Leslie, 1990). As Tyr25 is replaced by phenylalanine in CAT<sub>I</sub> it follows that the role of His195 is the more important in respect of both chloramphenicol and fusidate binding. Indeed, when Tyr25 of CAT<sub>III</sub> is replaced by phenylalanine it results in only minor changes in the kinetic parameters of the acetyl

transfer reaction (Murray *et al.*, 1991a). The observation that fusidate analogues wherein the 3- $\alpha$ -hydroxyl is replaced by a  $\beta$ -OH (3-epifusidate) or keto substituent (3-oxofusidate) bind weakly to wild-type CAT<sub>I</sub> (Bennett & Shaw, 1983) serves to emphasise the importance of the hydrogen bond with His195. Pre-steady state kinetic analyses (Day *et al.*, 1995) show that the rate of fusidate dissociation from wild-type CAT<sub>I</sub> increases when His195 is replaced by alanine but that the equivalent substitution in CAT<sub>III</sub> has no effect on inhibitor dissociation, implying that the hydrogen bond interaction is absent in wild-type CAT<sub>III</sub>. Indeed, one explanation of the requirement for the concerted substitution of four residues in the chloramphenicol binding site of CAT<sub>III</sub> is that they are necessary to permit access of the inhibitor to the base of the substrate binding pocket to form the hydrogen bond with His195.

NE2 of His195 acts as a general base to abstract a proton from the primary (C-3) hydroxyl of chloramphenicol as the initial step in the acetyl transfer reaction. Although the 3- $\alpha$ -hydroxyl of bound fusidate could in principle also donate its proton to His195 it is  $\sim 1.6$  Å removed from the position occupied by the primary (C-3) hydroxyl of chloramphenicol and in an entirely inappropriate orientation to attack the thioester carbonyl of acetyl-CoA and form a productive tetrahedral intermediate after proton abstraction (Figure 7). In addition, the C-2, C-3 and C-4 atoms of the A-ring of fusidate partly overlap with the positions of atoms of the tetrahedral intermediate (Lewendon *et al.*, 1990). This accounts for the inability of fusidate binding CAT variants to acetylate the steroidal inhibitor.

Using a sphere of radius 1.4 Å as probe we calculated that 575 Å<sup>2</sup> (or 81%) of the solvent-accessible area of fusidate becomes buried on binding to CAT. A similar proportion (89%, 1.7 Å probe radius) of progesterone is buried in its complex with the anti-progesterone monoclonal antibody DB3

Table 4. CAT-fusidate contacts less than 4 Å

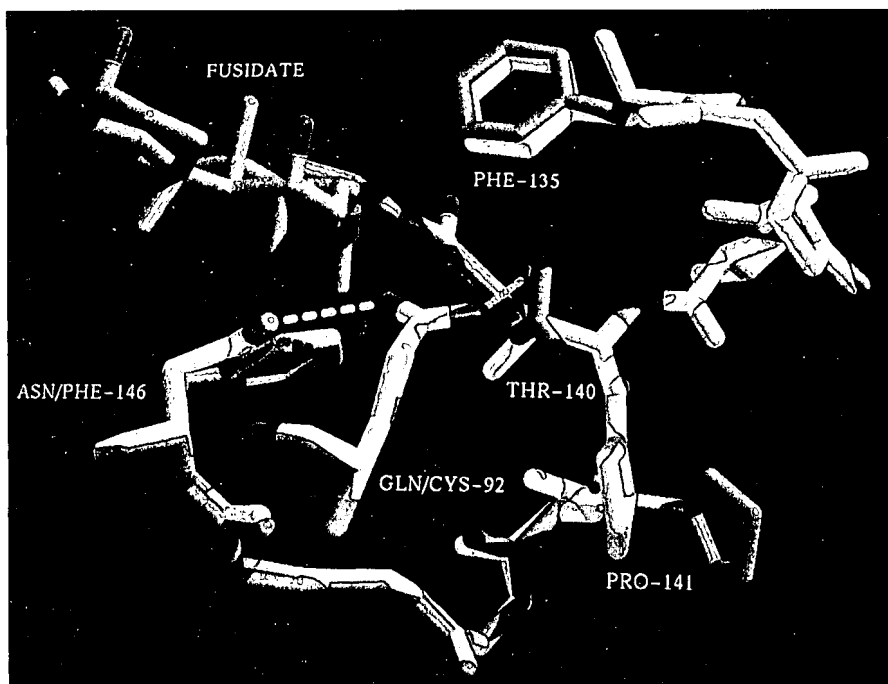
Fusidate atom	CAT atom	Distance (Å)
C1	OH Tyr25	3.59
	CG2 Thr94	3.96
C2	OH Tyr25	3.84
	CG2 Thr94	3.50
	CE1 Phe103	3.77
	CZ Phe103	3.93
C3	O Wat448	3.84
	OH Tyr25	3.82
	NE2 His195	3.49
	O Wat436	3.96
	O Wat437	3.83
C4	O Wat437	3.66
C9	CD2 Phe146	3.91
C11	CD2 Phe146	3.89
C12	CE2 Phe135	3.78
	CE2 Phe146	3.14
	CD2 Phe146	3.37
C13	CE2 Phe146	3.77
C15	CD2 Leu29	3.07
C17	CE2 Phe146	3.89
	CZ Phe146	3.98
	CD2 Leu160	3.17
C18	NE2 His195	3.43
	CD2 His195	3.48
	O Wat437	3.87
	CB Phe146	3.89
C19	CE1 Phe24	3.85
	OH Tyr25	3.47
	CE1 Tyr25	3.68
C20	CD2 Leu29	3.69
	CZ Phe146	3.08
	CE2 Phe146	3.10
	CE1 Phe146	3.29
	CD2 Phe146	3.38
	CD1 Phe146	3.53
	CG Phe146	3.57
C22	CZ Phe146	3.86
	CE2 Phe146	3.93
	CE2 Phe135	3.94
C23	CE2 Phe146	3.67
	CZ Phe146	3.90
	O Wat314	3.66
	O Wat314	3.68
C24	SG Cys92	3.88
C26	SG Cys92	3.66
C27	OG Ser107	3.69
	CD2 Phe135	3.04
C28	CE2 Phe135	3.50
	CG Phe135	3.82
	O Wat314	3.41
	CE2 Phe168	3.78
C31	CD2 Phe168	3.99
C32	CE2 Phe168	3.11
	CD2 Phe168	3.28
O1	OH Tyr25	3.54
	CE2 Phe135	3.86
	CZ Phe135	3.92
O2	CZ Phe146	3.90
O6	OH Tyr25	2.79 <sup>a</sup>
	CZ Tyr25	3.41
	CE1 Tyr25	3.93
	CZ Phe103	3.59
	NE2 His195	2.42 <sup>a</sup>
	CE1 His195	3.15
	CD2 His195	3.33
	O Wat436	3.73

<sup>a</sup> Hydrogen bond interactions.

1993). However, in the latter case it is probable that the eight-carbon hydrophobic tail of the substrate cholesterol (which is replaced by a keto group in dehydroandrosterone) projects out of the binding pocket in a manner strictly analogous to that seen in the CAT-fusidate complex. In each structure the steroid binding site is primarily formed from the side-chains of hydrophobic amino acid residues but the precise spectrum of contacts is quite different in each case. All three proteins utilise hydrogen bonds to histidyl residues to stabilise the oxygen atom (keto or hydroxyl) of the C-3 substituent of the A-ring. In cholesterol oxidase His447, which is hydrogen bonded to the 3-keto substituent *via* a bridging water molecule, is implicated in each of several postulated mechanisms for the oxidation reaction (Li *et al.*, 1993).

Tight binding of steroidal and other highly apolar molecules to proteins requires not only a high degree of surface and steric complementarity between ligand and binding site but also a mechanism to address the energetic cost of a hydrophobic binding pocket which is accessible to bulk solvent in the unliganded state. In the cases of progesterone binding to the DB3 Fab' (Arevalo *et al.*, 1993), and of dehydroandrosterone to cholesterol oxidase (Li *et al.*, 1993), desolvation is achieved by conformational changes in the protein which accompany ligand binding. In the unliganded state the binding site of DB3 exists in a "closed" conformation, such that its hydrophobic amino acids are shielded from bulk solvent, which can convert to an open state to permit access of the progesterone. In cholesterol oxidase, the substrate binding site is completely isolated from bulk solvent both in the absence (when the site is occupied by ordered water molecules) and in the presence of the steroid. Thus, a model for steroid binding to cholesterol oxidase requires that the ligand binding cavity first open to permit binding then reclose around the substrate, with the concomitant displacement of the ordered water. In contrast to the above examples, the binding of fusidate to CAT does not appear to be accompanied by conformational changes of the protein to facilitate its access to the active site cleft at each subunit interface. This implies that CAT<sub>III</sub> can tolerate the (presumably destabilising) effect of exposing three "new" hydrophobic residues to solvent in the unliganded state. In this respect it is perhaps significant that wild-type CAT<sub>III</sub> is an extremely robust enzyme, being highly resistant to thermal denaturation (Lewendon *et al.*, 1988) and remaining folded and trimeric in 8 M urea (P.J.D., unpublished experiments). Although the stability of the Q92C/N146F/Y168F/I172V variant has not been studied, it should be noted that CAT<sub>I</sub>, the chloramphenicol binding site of which is even more hydrophobic than that of the CAT<sub>III</sub> quadruple mutant, is both less thermostable and less soluble than wild-type CAT<sub>III</sub> (Day *et al.*, 1995).

(Arevalo *et al.*, 1993), whereas dehydroandrosterone is completely enclosed and sequestered from bulk solvent when bound to cholesterol oxidase (Li *et al.*,

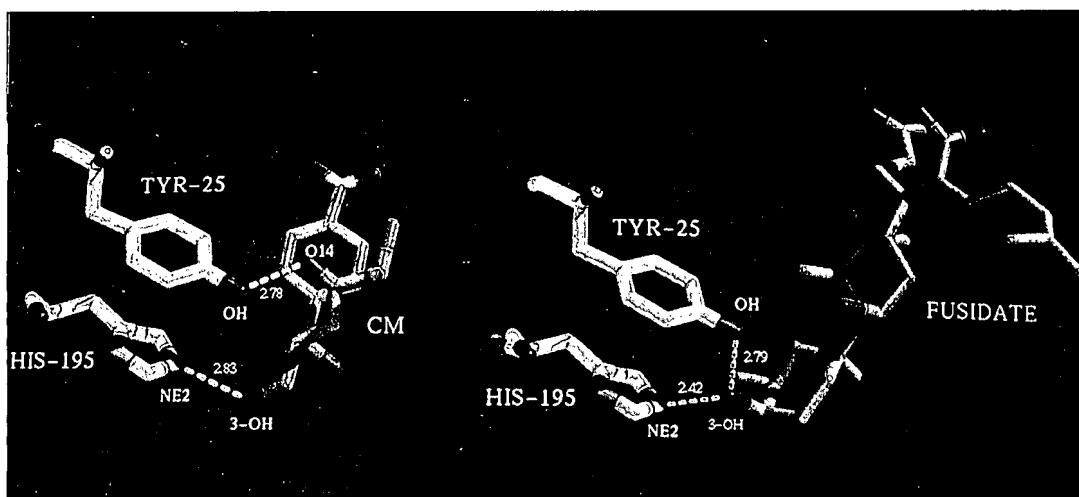


**Figure 5.** Surface loop movement accompanying the Q92C substitution. The structures of wild-type and the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub> were aligned by superposition of the main-chain atoms of residues 10 to 130 and 150 to 210. Residues 92 and 135 to 146 of the wild-type (white) and substituted (green) variants are shown in addition to the fusidate molecule bound to the latter. Note that the side-chains of residues 136–137 and 142–145 have been deleted from the image for the purpose of clarity. Replacement of Gln92 by cysteine eliminates the hydrogen bond to Thr140 permitting the movement of a surface loop (residues 137 to 142) which would otherwise preclude fusidate binding. Although Val139 and Thr140 are disordered in the CAT-fusidate complex and therefore not shown, the loop movement is revealed by a shift of several Å in the position occupied by Pro141.

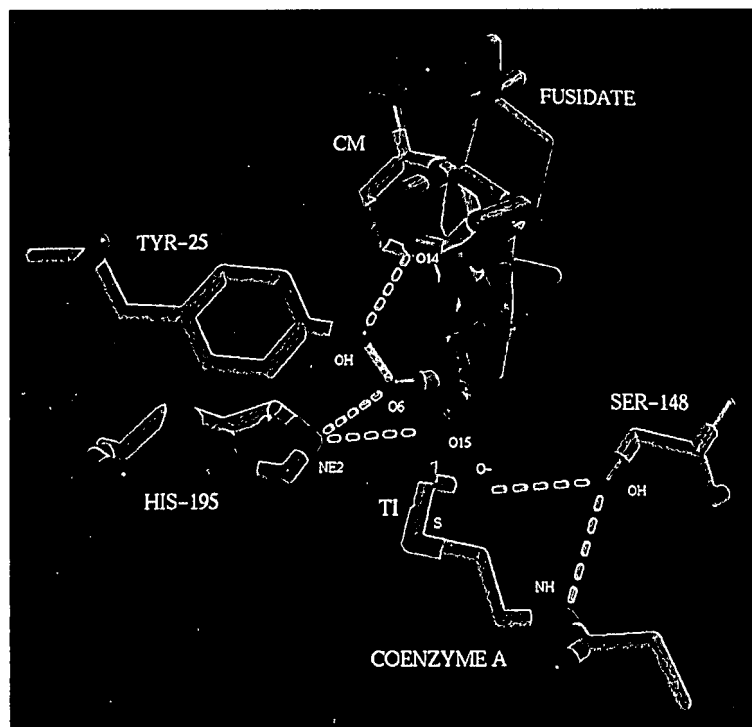
## Conclusions

We have used site-directed mutagenesis and X-ray crystallography to investigate the mechanism by which a single enzyme active site can bind two competing but chemically dissimilar ligands with approximately equal avidity. The structure of

fusidate bound to the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub> reveals how a single protein can confer resistance *in vivo* to two entirely different classes of antimicrobial agent and is the third example (after  $\beta$ -lactamase and CAT itself) of an antibiotic resistance mechanism which is understood at the atomic level. It has been suggested that



**Figure 6.** Intermolecular hydrogen bonds between bound ligand and the side-chains of Tyr25 and His195 occur in both CAT-chloramphenicol and CAT-fusidate complexes.



**Figure 7.** Superposition of bound fusidate inhibitor and the structure of the complex of wild-type CAT<sub>III</sub> and the tetrahedral oxyanion intermediate of the acetyl-transfer reaction. In the acetyl-transfer reaction catalysed by CAT, abstraction (by NE2 of His195) of the 3-hydroxyl (O-15) proton of chloramphenicol facilitates nucleophilic attack at the carbonyl carbon of acetyl-CoA leading to the formation of an oxyanion tetrahedral intermediate (O<sup>-</sup>) which is stabilised by a hydrogen-bond interaction with the side-chain of Ser148 (Lewendon, *et al.*, 1990). While, in principle, proton abstraction from the 3- $\alpha$ -hydroxyl (O-6) substituent of fusidate may also occur, the position and orientation of this hydroxyl group (and steric hindrance due to overlap with the C-2 and C-3 carbon atoms of the A-ring) preclude formation of a productive tetrahedral intermediate, accounting for the observed failure of CAT to acetylate the steroidal inhibitor. For the purpose of clarity only the A, B and C rings of fusidate and those parts of CoA immediately proximal to the tetrahedral intermediate (TI) are shown.

co-administration of chloramphenicol and fusidate (or fusidate analogues) might offer a possible route to overcome CAT-mediated chloramphenicol resistance in clinical practice (Davies, 1994). Our structural data suggest that such a strategy may not be promising, since, of the diverse range of naturally occurring CAT variants, it is only CAT<sub>I</sub> that carries the requisite motif of amino acid residues in the chloramphenicol binding site to generate a high affinity fusidate site. Conversely, it is apparent that mutations leading to the substitution of amino acids within the chloramphenicol binding site of a fusidate-binding CAT variant (c.f. CAT<sub>I</sub>) might be expected to result in loss of inhibitor affinity without serious impairment of catalytic competence in the acetyl transfer reaction or consequent chloramphenicol resistance. It is not apparent whether CAT<sub>I</sub>-mediated fusidate resistance is an evolved phenomenon or merely the result of a serendipitous arrangement of side-chains at the active site of this one variant. As only four of the 17 residues of the chloramphenicol binding site are conserved among known CAT variants it has not been possible to infer the evolutionary relationships between members of the family. Nonetheless, in the context of fusidate resistance, it is probably

significant that CAT<sub>I</sub> is of enterobacterial origin (including numerous genera that are relatively insensitive to fusidate due to limited outer membrane permeability) whereas naturally occurring CAT variants isolated from Gram-positive genera (commonly fusidate-sensitive) do not, to the best of our knowledge, confer resistance.

The CAT-fusidate complex is the third example of the determination of the structure of a protein bound to a steroidal ligand and one wherein the architecture of the binding site and the likely mechanism of ligand binding are both quite distinct from those observed in previously published structures. Because the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub> is only slightly compromised in the acetyl-transfer reaction (Table 2B) we believe that the active site of the protein is not significantly deformed in the absence of bound fusidate. This implies that the hydrophobic residues of the fusidate binding site are solvent-exposed in the unliganded enzyme and that, in contrast to cholesterol oxidase and the anti-progesterone Fab' DB3, significant conformational changes are not required to facilitate steroid binding. It is not known whether the determinants and dynamics of ligand binding to the effector recognition domains of

steroid receptors are analogous to those exemplified by the structures of the CAT-fusidate complex (binding at a preformed cleft), to that of DB3 and progesterone (opening of a closed binding site), or to the cholesterol oxidase-dehydroandrosterone complex (gating and reclosure of a preformed cavity). However, it is probable that the several discrete but linked functions of receptors of the steroid/vitaminD/thyroid hormone superfamily (Evans, 1988) will involve additional structural responses to ligand binding, favouring dimerisation and modulating the specificity and affinity of receptor interaction with DNA response elements.

## Materials and Methods

### Construction of recombinant genes and site-directed mutagenesis

Regions of the DNAs encoding CAT<sub>I</sub> (Alton & Vapnek, 1979) and CAT<sub>III</sub> (Murray *et al.*, 1988) were recombined *in vitro* using the technique of "sticky-feet" directed mutagenesis (Clackson & Winter, 1989). Single and multiple point mutations of CAT<sub>III</sub> were introduced using oligonucleotide primers and single-stranded M13 DNA templates loaded with deoxyuridine by preparation in the *dut ung E. coli* strain RZ1032 (Künkel *et al.*, 1987). The presence of desired mutations and the absence of second site changes were confirmed by DNA sequence determination of both the complete coding and the 5' non-coding regions of the genes.

### Expression and purification of CAT

Wild-type and mutant CAT proteins were expressed in *E. coli* JM101 after subcloning the coding sequences in pUC18. Enzymes were purified from cell-free extracts by affinity chromatography using chloramphenicol-Sepharose (Lewendon *et al.*, 1988) or by ion-exchange chromatography (DEAE-Sepharcel) followed by affinity chromatography using Cibacron-blue Sepharose (Murray *et al.*, 1991a). Homogeneity of purified enzymes was confirmed by SDS-polyacrylamide gel electrophoresis and enzyme concentrations were determined by the method of Lowry *et al.* (1951).

### Assay of CAT activity and kinetic determinations

CAT activity was assayed spectrophotometrically at 25°C as described previously (Lewendon *et al.*, 1990). Standard assays contained 0.4 mM acetyl-CoA (prepared from the lithium salt of CoA, Pharmacia, by the method of Simon & Shemin, 1953), 0.1 mM chloramphenicol and 1.0 mM 5,5'-dithiobis(2-nitrobenzoic acid) in TSE buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA). One unit is defined as the amount of enzyme required to convert 1 µmol of chloramphenicol to 3-acetylchloramphenicol in one minute using the standard assay. Concentrations of acetyl-CoA and chloramphenicol were varied in the standard assay for determination of steady state kinetic parameters and all assays were carried out in triplicate. Initial rate values were used to construct double reciprocal plots and kinetic parameters were derived from slope and intercept replots (Kleanthous & Shaw, 1984).  $K_i$  values for competitive inhibition by sodium fusidate were determined by varying the

concentrations of chloramphenicol and inhibitor in standard assays containing a fixed concentration of acetyl-CoA (routinely  $\sim 5 \times K_m$ ), and were calculated from linear slope replots derived from double reciprocal plots. Crude extracts of chimaeric enzymes which appeared to be inactive in the standard assay were screened for low levels of CAT activity using a sensitive radiometric assay (Gorman *et al.*, 1982).

### Crystallisation and structure determination

Single crystals of the Q92C/N146F/Y168F/I172V variant of CAT<sub>III</sub> were prepared by microdialysis using small "Lucite" buttons as described previously (Leslie, 1990). Each button contained 25 µl of protein ( $\sim 5.7$  mg ml<sup>-1</sup>) in 10 mM MES buffer (pH 6.3), or the same buffer supplemented with 0.5 mM sodium fusidate, and were dialysed at 4°C against 10 ml of 10 mM MES (pH 6.3) containing 2 to 4% (v/v) 2-methyl-2,4-pentane-diol, 1 mM hexamine cobalt (III) chloride, 0.1 mM dithiothreitol and either 1 mM chloramphenicol or 0.5 mM sodium fusidate. Crystals were isomorphous with those of the wild-type enzyme; space group *R*32 (equivalent hexagonal cell dimensions  $a = 107.8$  Å,  $c = 124.1$  Å). 80° of data were collected to 2.2 Å resolution from a single crystal (dimensions 260 µm × 240 µm × 120 µm) grown in the presence of fusidate using CuKα radiation from a GX13 rotating anode generator with double mirror collimation. Data were recorded on a prototype Hendrix-Lentfer image plate scanner with a diameter of 18 cm. The images were integrated with MOSFLM (Leslie, 1992) and programs from the CCP4 Suite (1994). A total of 68,511 observations were reduced to a unique dataset of 14,238 reflections with a crystallographic merging *R*-factor of 9.5% (24.9% in the highest resolution range). The dataset is 99% complete out to 2.2 Å resolution, with an overall  $I/\sigma(I)$  ratio of 18.3 (7.6 at 2.2 Å resolution).

The refined structure of the CAT<sub>III</sub>-chloramphenicol binary complex (Leslie, 1990) was used as a starting model for refinement. The chloramphenicol and all water molecules in the chloramphenicol binding pocket were removed from the model and the substitutions Q92C, N146F, Y168F and I172V made. This model was subjected to alternating rounds of refinement using the CCP4 programs SFALL, PROTIN and PROLSQ and manual rebuilding using the interactive graphics program O (Jones *et al.*, 1991). After one round of refinement there was very clear density compatible with a model of fusidate derived from the crystal structure of fusidic acid methyl ester *p*-bromobenzoate (Cooper & Hodgkin, 1968) and this was included in the model for all subsequent refinement. After four rounds of model building and refinement the final *R*-factor was 17.4% for all reflections in the resolution range 6.0 to 2.2 Å. Computer graphics images were produced using the *conic* option (Huang *et al.*, 1991) of the MidasPlus program (Ferrin *et al.*, 1988).

## Acknowledgements

This work was supported by project grants from the Medical Research Council and Science and Engineering Research Council. M.J.S. is a Royal Society University Research Fellow. Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081).

## References

- Alton, N. K. & Vapnek, D. (1979). Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. *Nature*, **282**, 864–869.
- Arevalo, J. H., Stura, E. A., Taussig, M. J. & Wilson, I. A. (1993). Three-dimension structure of an anti-steroid Fab' and progesterone-Fab' complex. *J. Mol. Biol.* **231**, 103–118.
- Bennett, A. D. & Shaw, W. V. (1983). Resistance to fusidic acid in *Escherichia coli* mediated by the type I variant of chloramphenicol acetyltransferase. A plasmid-encoded mechanism involving antibiotic binding. *Biochem. J.* **215**, 29–38.
- Betz, J. L. & Sadler, J. R. (1981). Variants of a cloned synthetic lactose operator. II. Chloramphenicol-resistant revertants retaining a lactose operator in the CAT gene of plasmid pBR325. *Gene*, **15**, 187–200.
- Clackson, T. & Winter, G. (1989). 'Sticky-feet'-directed mutagenesis and its application to swapping antibody domains. *Nucl. Acid. Res.* **17**, 10163–10170.
- Collaborative Computational Project, Number 4 (1994). *Acta Crystallog. sect. D*, **50**, 760–763.
- Cooper, A. & Hodgkin, D. C. (1968). The crystal structure and absolute conformation of fusidic acid methyl ester 3-p-bromobenzoate. *Tetrahedron*, **24**, 909–922.
- Cundliffe, E. (1981). In *Molecular Basis of Antibiotic Action* (Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. & Waring, M. J., eds), pp. 405–547, Wiley Interscience, London.
- Davies, J. (1994). Inactivation of antibiotics and the dissemination of resistance genes. *Science*, **264**, 375–382.
- Day, P. J., Murray, I. A. & Shaw, W. V. (1995). Properties of hybrid active sites in oligomeric proteins: kinetic and ligand binding studies with chloramphenicol acetyltransferase trimers. *Biochemistry*, **34**, 6416–6422.
- Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. *Science*, **240**, 889–895.
- Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graphics*, **6**, 13–27.
- Ghosh, D., Weeks, C. M., Grockhulski, P., Duax, W. L., Erman, M., Rimsay, R. L. & Orr, J. C. (1991). Three-dimensional structure of holo- $3\alpha,20\beta$ -hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. *Proc. Natl Acad. Sci. USA*, **88**, 10064–10068.
- Gorman, C. M., Moffat, L. & Howard, B. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**, 1044–1051.
- Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. (1991). Conic: a fast renderer for space-filling molecules with shadows. *J. Mol. Graphics*, **9**, 230–236.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
- Kleanthous, C. & Shaw, W. V. (1984). Analysis of the mechanism of chloramphenicol acetyltransferase by steady-state kinetics. Evidence for a ternary complex mechanism. *Biochem. J.* **223**, 211–220.
- Kleanthous, C., Cullis, P. M. & Shaw, W. V. (1985). 3-(Bromoacetyl) chloramphenicol, an active site directed inhibitor of chloramphenicol acetyltransferase. *Biochemistry*, **24**, 5307–5313.
- Künkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.
- Leslie, A. G. W. (1990). Refined crystal structure of chloramphenicol acetyltransferase at 1.75 Å resolution. *J. Mol. Biol.* **213**, 167–186.
- Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing films and image plate data. In CCP4 and ESF-EACMB Newsletter on Protein Crystallography, Number 26.
- Lewendon, A., Murray, I. A., Kleanthous, C., Cullis, P. M. & Shaw, W. V. (1988). Substitutions in the active site of chloramphenicol acetyltransferase: role of a conserved aspartate. *Biochemistry*, **27**, 7385–7390.
- Lewendon, A., Murray, I. A., Shaw, W. V., Gibbs, M. R. & Leslie, A. G. W. (1990). Evidence for transition-state stabilisation by serine-148 in the catalytic mechanism of chloramphenicol acetyltransferase. *Biochemistry*, **29**, 2075–2080.
- Lewendon, A., Murray, I. A., Shaw, W. V., Gibbs, M. R. & Leslie, A. G. W. (1994). Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: evidence for a general base role for glutamate. *Biochemistry*, **33**, 1944–1950.
- Li, J., Vrielink, A., Brick, P. & Blow, D. M. (1993). Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. *Biochemistry*, **32**, 11507–11515.
- Lowry, O. H., Rosebrough, N. H., Farr, A. L. & Randall, R. J. (1951). Protein measurement with the folin phenol reagent. *J. Biol. Chem.* **193**, 265–275.
- Mattevi, A., Obmolova, G., Schulze, E., Kalk, K. H., Westphal, A. H., de Kok, A. & Hol, W. G. (1992). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science*, **255**, 1544–1550.
- Morize, I., Surcouf, E., Vaney, M. C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E. & Mornon, J. P. (1987). Refinement of the C222<sub>1</sub> crystal form of oxidised uteroglobin at 1.34 Å resolution. *J. Mol. Biol.* **194**, 725–739.
- Murray, I. A., Hawkins, A. R., Keyte, J. W. & Shaw, W. V. (1988). Nucleotide sequence analysis and overexpression of the gene encoding a type III chloramphenicol acetyltransferase. *Biochem. J.* **252**, 173–179.
- Murray, I. A., Lewendon, A. & Shaw, W. V. (1991a). Stabilisation of the imidazole ring of His-195 at the active site of chloramphenicol acetyltransferase. *J. Biol. Chem.* **266**, 11695–11698.
- Murray, I. A., Lewendon, A., Williams, J. A., Cullis, P. M., Lashford, A. G. & Shaw, W. V. (1991b). A novel substrate for assays of gene expression using chloramphenicol acetyltransferase. *Nucl. Acids Res.* **19**, 6648.
- Packman, L. C. & Shaw, W. V. (1981). The use of naturally occurring hybrid variants of chloramphenicol acetyltransferase to investigate subunit contacts. *Biochem. J.* **193**, 541–552.
- Parker, M. G. (1993). *Steroid Hormone Action*, Oxford University Press, Oxford, U.K.
- Shaw, W. V. (1967). The enzymatic acetylation of chloramphenicol by extracts of *R*-factor resistant *Escherichia coli*. *J. Biol. Chem.* **242**, 687–693.
- Shaw, W. V. & Leslie, A. G. W. (1991). Chloramphenicol acetyltransferase. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 363–386.
- Simon, E. J. & Shemin, D. (1953). The preparation of S-succinyl coenzyme A. *J. Am. Chem. Soc.* **75**, 2520–2524.

- Westbrook, E. M., Piro, O. E. & Sigler, P. B. (1984). The 6-Å crystal structure of  $\Delta^5$ -3-ketosteroid isomerase. Architecture and location of the active center. *J. Biol. Chem.* **259**, 9096-9103.
- Zhao, J. & Aoki, T. (1992). Cloning and nucleotide sequence analysis of a chloramphenicol acetyltransferase gene from *Vibrio anguillarum*. *Microbiol. Immunol.* **36**, 695-705.

*Edited by A. R. Fersht*

*(Received 31 July 1995; accepted 26 September 1995)*



# Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from *Escherichia coli* and mammalian cells

Weiying Pan<sup>†</sup>, Elisabetta Ravot<sup>§</sup>, Ralf Tolle<sup>¶</sup>, Rainer Frank, Raphael Mosbach, Ivana Türbachova and Hermann Bujard<sup>\*</sup>

ZMBH, Universität Heidelberg, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

Received September 29, 1998; Revised and Accepted December 17, 1998

DDBJ/EMBL/GenBank accession no. AJ131294

## ABSTRACT

The *Plasmodium falciparum* malaria parasite is the causative agent of malaria tropica. Merozoites, one of the extracellular developmental stages of this parasite, expose at their surface the merozoite surface protein-1 complex (MSP-1), which results from the proteolytic processing of a 190–200 kDa precursor. MSP-1 is highly immunogenic in humans and numerous studies suggest that this protein is an effective target for a protective immune response. Although its function is unknown, there are indications that it may play a role during invasion of erythrocytes by merozoites. The parasite-derived *msh-1* gene, which is ~5000 bp long, contains 74% AT. This high AT content has prevented stable cloning of the full-size gene in *Escherichia coli* and consequently its expression in heterologous systems. Here, we describe the synthesis of a 4917 bp gene encoding MSP-1 from the FCB-1 strain of *P.falciparum* adjusted for human codon preferences. The synthetic *msh-1* gene (55% AT) was cloned, maintained and expressed in its entirety in *E.coli* as well as in CHO and HeLa cells. The purified protein is soluble and appears to possess native conformation because it reacts with a panel of mAbs specific for conformational epitopes. The strategy we used for synthesizing the full-length *msh-1* gene was to assemble it from DNA fragments encoding all of the major proteolytic fragments normally generated at the parasite's surface. Thus, after subcloning we also obtained each of these MSP-1 processing products as hexahistidine fusion proteins in *E.coli* and isolated them by affinity chromatography on Ni<sup>2+</sup> agarose. The availability of defined preparations of MSP-1 and its major processing products open up new possibilities for in-depth studies at the structural and functional level of this important protein, including the exploration of MSP-1-based experimental vaccines.

## INTRODUCTION

Malaria caused by *Plasmodium falciparum* infections continues to be a serious health problem in major parts of the world. The identification of targets for immunologic interventions against this infectious disease remains, therefore, an important goal. Merozoites, which are the extracellular developmental form of the parasite that invade erythrocytes, expose on their surface a protein complex, which is the processing product of a 190–200 kDa precursor known as merozoite surface protein-1 (MSP-1; 1–5). Upon deposition in the merozoite membrane via a glycosylphosphatidylinositol (GPI) anchor, this precursor is proteolytically cleaved during late schizogony into fragments which remain associated with the parasite surface. Sequence analysis of *msh-1* genes of different *P.falciparum* strains has revealed that major regions of the protein are dimorphic belonging to either the K1 or the MAD20 prototype (Fig. 1A) while other parts are highly conserved (3). Two small regions near the N-terminus show higher variability. These features, as well as the presence of point mutations scattered throughout the molecule and evidence for intragenic recombination and/or gene conversion, confer a surprisingly limited polymorphism to these abundant surface proteins (3,5).

A number of experimental findings suggest that MSP-1 of *P.falciparum* may elicit a protective immune response against infections by the parasite. For example, in the rodent model, immunization of mice with the analogous protein of *Plasmodium yoelii* yielded protection (6), as did the transfer of monoclonal antibodies and immune serum against this protein (7–9). Sero-epidemiologic data (10,11) and results from several vaccination trials conducted with various *P.falciparum*-derived MSP-1 preparations in non-human primates also support the candidacy of this protein or parts thereof as promising components of a subunit vaccine against malaria tropica. In these trials, *Saimiri* or *Aotus* monkeys were immunized either with MSP-1 isolated from parasites (12–15) or with synthetic peptides or recombinant MSP-1 fragments (15–21). The recombinant fragments assessed most recently were primarily from the C-terminal region of MSP-1. Although the data from these trials support MSP-1 as a vaccine

\*To whom correspondence should be addressed. Tel: +49 6221 548214; Fax: +49 6221 545892; Email: bujardh@sun0.urz.uni-heidelberg.de

Present addresses: <sup>†</sup>Second Military Medical University, Department of Parasitology, 800 Xiang Yin Road, Shanghai 200433, China, <sup>§</sup>RIGHT, IRCC Policlinico San Matteo, Padiglione Forlanini, P. le Golgi 2, I-27100 Pavia, Italy and <sup>¶</sup>LION Bioscience AG, INF 515, D-69120 Heidelberg, Germany

candidate, the MSP-1 protective effects measured in these trials barely satisfy requirements for statistical significance because group sizes were generally too small and since collected data were not confirmed in strictly comparable, subsequent trials. The two main reasons for this situation are the scarceness of suitable experimental animals and the difficulties associated with preparing sufficient amounts of well-defined MSP-1 from parasites. On the other hand, expression of full-length recombinant MSP-1 in heterologous systems has turned out to be most difficult if not impossible (22,23). This appeared primarily due to the high AT content of *P.falciparum* DNA which prevents the cloning and stable maintenance of large genes in *Escherichia coli*, thereby precluding crucial studies at the genetic and biochemical level which may have led to the elucidation of its function.

Thus, we decided to synthesize a 4917 bp polynucleotide encoding MSP-1 of the Colombia FCB-1 strain (24,25) and change the AT content such that it could be maintained and expressed in a variety of hosts. Herein, we describe the design and synthesis of this gene based on human codon frequencies. We also report the cloning of this synthetic gene and its controlled expression in *E.coli* and in mammalian cells. Purification of full-length protein from both expression systems yielded material that is recognized by several monoclonal antibodies known to interact with conformational epitopes. Moreover, the strategy employed for the synthesis of the gene allowed for subcloning, production and purification of all major processing products of MSP-1. The availability of large amounts of MSP-1 and its fragments opens up new possibilities for the thorough investigation of this prominent malaria antigen.

## MATERIALS AND METHODS

### Sequence design

The amino acid sequence of MSP-1 of the Colombia FCB-1 strain (25) was translated into a DNA sequence with an average codon composition similar to that found in human coding sequences (26).

This was achieved by using a random number generator to make each codon assignment, a process that proved to be truly random, because each run of the program yielded a different synthetic allele of MSP-1. One sequence was chosen as the master sequence and modified in a number of ways to eliminate sequences that might be detrimental to efficient transcription and translation of the synthetic gene. All analysis programs mentioned below were from the Genetics Computer Group program collection (27). Positions where the introduction of additional endonuclease cleavage sites appeared feasible without changing the amino acid sequence were identified with the 'Map' program using the option 'silent'. 'Find Patterns' was used to search for consensus sequences that are indicative of prokaryotic promoters, poly(A) signals and exon-intron boundaries. Prokaryotic factor-independent RNA polymerase terminator structures were identified with the 'Terminator' program. Inverted repeats which might lead to the formation of undesirable secondary structures were identified with the 'Stemloop' program. All these structures, when encountered, were eliminated by using alternative codons. Moreover, long runs of purines (>7 nt) that may cause transcriptional termination in some viral systems were disrupted. Finally, the stability of the resulting RNA molecule was assessed with 'Fold RNA'. This analysis was performed on overlapping fragments because the software restricted the length of the input sequence to 1200 bases. Any structures more stable than the mRNA of the human glyceraldehyde-3-phosphate dehydrogenase were eliminated.

### Design and synthesis of oligonucleotides

The oligonucleotide primers used for the PCR-based synthesis of high molecular weight, double-stranded DNA were restricted in size to maximally 120 nt in order to ensure that at least 50% of the PCR products were error free. The primers were designed using the Oligo 4.0 program and the following parameters were considered. The overlapping region between two oligonucleotides used in one PCR reaction was on average 18 nt long. The internal stability of each overlapping oligo pair was  $-\Delta G \geq 9$  kcal/mol. Potential hairpin and duplex structures with  $-\Delta G \geq 8$  kcal/mol were eliminated as were false priming sites. The oligonucleotides were synthesized on a 1000 Å pore size glass support using an Applied Biosystems 394 synthesizer following standard protocols. They were purified by electrophoresis in 10% polyacrylamide (PAA) containing 7 M urea. Upon electroelution from gels, they were precipitated with ethanol.

### Asymmetric PCR-based synthesis of double-stranded DNA fragments

The overall strategy of fragment synthesis is outlined in Figure 2. In general, four assays were performed in parallel. The ratio of the oligonucleotide pair used in each reaction was 5:1 in order to yield the appropriate asymmetrically amplified product (28). Products A, B, C and D, respectively, resulted from four PCR assays that contained the oligonucleotide pairs O1/O2 and O5/O6 in ratios of 25:5, O3/O4 and O7/O8 in ratios of 5:25 pmol, respectively. Five PCR cycles were performed in 50 µl of 10 mM Tris-HCl, 1.5 mM MgCl<sub>2</sub>, pH 8.3, containing 2.5 U Taq polymerase (Boehringer, Mannheim) and the four deoxynucleotide triphosphates (200 µM each). The optimized cycling conditions (Omnigene TR3 CM220 thermocycler) were 10 s at 94°C (denaturation), 30 s at 55°C (annealing) and 60 s at 72°C (polymerization). Product E was prepared by combining the products from assays A and B, and F from assays C and D. After amplifying for five cycles, we added 25 pmol of O1 and O8 to 5 pmol of E and F, respectively, to re-establish asymmetric oligonucleotide compositions and amplified for an additional eight cycles. Product G was prepared by combining these latter assays and amplifying for another 12 cycles. PAGE was used to follow the various synthetic steps. The final product G was separated from other reaction products by electrophoresis in 1% agarose gels, eluted from the gel slice according to the Qiaex II procedure (Qiagen, Düsseldorf), digested with *Bam*HI and *Cl*al and ligated into an appropriately cleaved pBSK\* plasmid. This plasmid is identical with pBSK (Stratagene, Heidelberg) except that the *Xba*I and *Spe*I sites within the multiple cloning site were replaced with *Nhe*I, *Mlu*I, *Nco*I and *Syl*I sites. The resulting vectors were transferred into *E.coli* strain SG13009. DNA fragments of the expected size liberated from plasmids of bacterial clones after *Bam*HI/*Cl*al cleavage were further analysed. Usually, 10–20 clones containing such inserts were sufficient to identify either error-free 600–800 bp DNA fragments or fragments containing small numbers of errors that could be eliminated by combining the error-free portions of two fragments via an internal cleavage site. All fragments were finally combined via their compatible unique cleavage sites positioned at either end (Fig. 1B). After assembly was complete, the *m*sp-1<sup>SI</sup> gene was sequenced in its entirety with a standard set of sequencing primers.

**Synthesis of full size MSP-1 and of MSP-1 fragments in *E. coli***

The DNA encoding MSP-1<sup>S2</sup>, p83 (without signal peptide, Fig. 1B), p30, p38, p42 and p19 (the latter without GPI anchor signal, Fig. 1B), respectively, were transferred from pBSK\* vectors to the expression vector pDS56 (29) via their *Bam*HI and *Cla*I cleavage sites. By utilizing the *Bam*HI insertion site of this vector, six histidines were attached to the N-terminus of the respective protein. This allows the purification of the resulting protein via Ni<sup>2+</sup> chelate chromatography (29).

The various vector constructs were transferred into the Lac repressor providing *E. coli* SG13009. Cultures derived from individual clones containing the proper plasmids were grown to early log phase (OD<sub>600</sub> 0.2) and induced with IPTG (1 mM) for 3 h. With the exception of p38, all fragments as well as the full-size p190 were produced in high yields (between 2 and 10% of the total protein), but even p38 was readily purified in sufficient quantity to allow characterization. For isolation of the various expression products, cell pellets were dissolved in 6 M guanidinium hydrochloride and applied directly to Ni<sup>2+</sup> chelate columns. Before the adsorbed material was eluted, the column was developed with a reverse gradient running from 6 to 1 M urea in 0.5 M NaCl, 0.05 M Tris-HCl, 20% glycerol, pH 7.4. Elution with an imidazole gradient (0–500 mM imidazole hydrochloride in 0.05 M Tris-HCl, 10% glycerol, pH 7) yielded MSP-1<sup>S2</sup> as well as all the fragments in highly purified and soluble form.

**Expression of MSP-1-encoding DNA in mammalian cells**

The *msp-1*<sup>S1</sup> gene was inserted as a *Mlu*I-*Cla*I fragment into plasmid pBi-5 (30) where it is co-regulated with the luciferase gene by the bidirectional promoter P<sub>bi-1</sub>. The activity of P<sub>bi-1</sub> is entirely dependent on tTA, the tetracycline controlled transcriptional activator (31). The resulting plasmid pBi-5/MSP-1<sup>S1</sup> was used to transiently express the gene in HeLa and CHO cells that constitutively produce tTA. Thus, HtTA-1 (31) and CtTA-1 cells were co-transfected with pBi-5/MSP-1<sup>S1</sup> and pUHD16-1 following a modified (30) calcium phosphate method. The latter plasmid gives rise to  $\beta$ -galactosidase which serves as a standard for determining transfection efficiencies. MSP-1<sup>S1</sup> expression was induced by removal of doxycycline (Dox) from the culture and cells were harvested after 30 h to determine luciferase activities in cell extracts as described previously (31). The production of MSP-1 was visualized by western blot analysis using monoclonal antibody mAb 5.2 (32) and compared with lysates prepared from uninduced cells.

**HtTA-1 and CtTA-1 cell lines that control the synthesis of MSP-1<sup>S1</sup>**

To integrate the *msp-1*<sup>S1</sup> gene controlled by P<sub>bi-1</sub> into the genome of HtTA-1 cells, the cells were grown in 35 mm dishes to 40–50% confluency and co-transfected with 2.9  $\mu$ g of linearized plasmid pBi-5/MSP-1<sup>S1</sup> and 0.1  $\mu$ g of linearized plasmid pHMR272 (30), which confers resistance to hygromycin B. After 24 h, the cells were transferred to 10 cm dishes and maintained in medium containing 300  $\mu$ g/ml hygromycin B. Resistant clones were isolated, expanded separately and analysed for luciferase activity (31) in the absence and presence of Dox (100 ng/ml). Several clones which exhibited efficient regulation of luciferase activity in a Dox-restricted manner were then analysed for tTA-dependent production of MSP-1 by western blot analysis. Further subcloning produced the cell line HtTA-9319 which efficiently expressed

MSP-1 as well as the *luc* gene. The MSP-1-expressing CHO cell line CtTA-27/29 was generated in an analogous way.

**Purification of MSP-1<sup>S1</sup> from HeLa cells by immunoaffinity chromatography**

HtTA-93/9 cells were grown to confluency in 10 cm dishes containing EMEM medium supplemented with 10% FCS. Cells from 20 such cultures were washed twice with PBS and suspended in 4 ml TNET lysis buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, 1% Triton X-100, pH 7.4) containing a cocktail of protease inhibitors (PMSF, aprotinin, antipain, bestatin, pepstatin and leupeptin, each at 5  $\mu$ g/ml, and 50  $\mu$ g/ml TLCK). The suspension was kept in ice for 30 min before it was centrifuged at 300 000 g and 4°C for 30 min. The supernatant was 'cleared' by passing it through a 1.5 ml column packed with protein A-Sepharose 4 fast flow (Pharmacia Biotech) and the flow-through was collected. For immunoaffinity chromatography the flow-through was then applied to a 1.5 ml mAb 5.2 (ATCC, HB9148)/protein A-Sepharose 4 fast flow column (33) equilibrated with TNET buffer. The column was washed with 5 bed vol of TNET, pH 7.4, followed by 5 bed vol of TNET, 0.65 M NaCl, pH 8.0, and by 2 bed vol of 150 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 6.8. Adsorbed protein was eluted with 0.1 M glycine buffer, pH 2.5, and fractions were collected into a previously titrated volume of 1 M Tris-HCl, pH 8.0. The protein was stored in 20% glycerol at -20°C.

**Analysis of MSP-1 isolated from *E. coli* and CHO cells by western blot**

MSP-1<sup>S</sup> purified from *E. coli* via Ni<sup>2+</sup> chelate chromatography and MSP-1<sup>S1</sup> isolated from CHTA-27/29 cells by immunoaffinity chromatography was subjected to electrophoresis in 8% PAA in the presence of SDS (2%) but under non-reducing conditions. Transfer of the protein onto ImmobilonP membranes (Millipore) was carried out in transfer buffer (0.01% SDS, 25 mM Tris, 192 mM glycine, 20% methanol) for 90 min at 350 mA.

Approximately 0.1  $\mu$ g of protein was applied per lane. The membrane-bound proteins were exposed to the various monoclonal antibodies and visualized via anti-mouse IgG AP-conjugate (Sigma, A2179) following standard procedures (33).

**RESULTS****Design of a polydeoxyribonucleotide encoding an MSP-1 sequence**

The MSP-1 coding sequence chosen for redesign is from *Plasmodium falciparum* strain FCB-1. Table 1 shows the bias towards A and T in the codons of the parasite gene in comparison with codon frequencies found in human coding sequences. By back-translating the amino acid sequence of the *msp-1* gene into DNA with human codon frequencies, the AT content was reduced from 74 to 55%. The redesigned gene was further modified to exclude sequences that may be problematic during synthesis or cloning and expression of the polynucleotide in various heterologous systems (Materials and Methods). These included a perfect *E. coli* promoter sequence lying upstream of a consensus-type translational start signal which gave rise to efficient expression from an internal start site. By making use of the degeneracy of the genetic code, we eliminated all the potentially problematic sequences by changing individual base pairs without affecting the encoded amino acid sequence.

**Table 1.** Comparing the codon frequencies of the *msp-1* gene of *P.falciparum* with human coding frequencies reveals an extreme bias towards A/T-containing codons in the parasite DNA

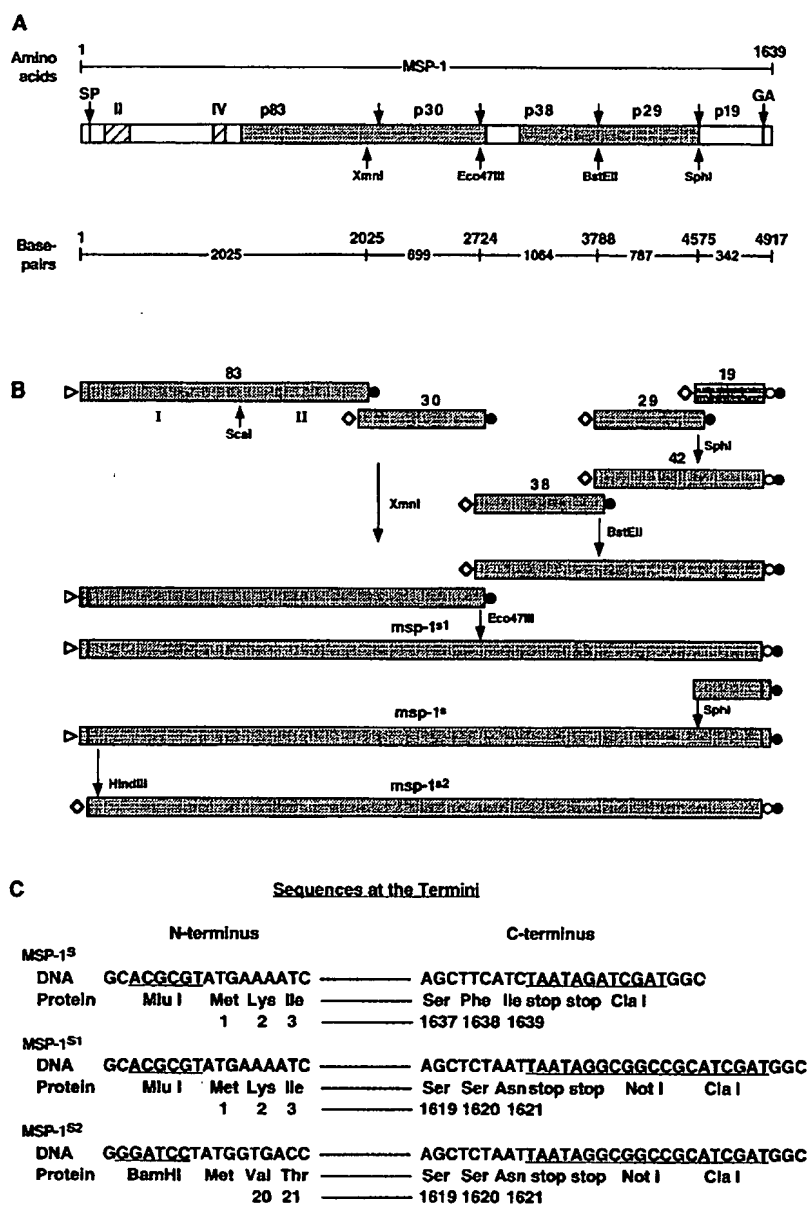
Codon frequencies (%)									
Amino acid	Codon	MSP-1 native	MSP-1 synthetic	human coding seq.	Amino acid	Codon	MSP-1 native	MSP-1 synthetic	human coding seq.
Ala	GCA	53	22	23	Leu	CTA	2	2	7
	GCC	8	49	40		CTC	3	30	20
	GCG	2	0	10		CTG	0	51	41
	GCT	37	29	27		CTT	17	8	13
Arg	AGA	62	30	20	Tyr	TTA	69	1	7
	AGG	10	17	20		TTG	9	8	13
	CGA	10	9	11	Lys	AAA	86	35	42
	CGC	0	26	19		AAG	14	65	58
	CGG	0	9	21	Met	ATG	100	100	100
	CGT	19	9	9					
Asn	AAC	22	58	54	Phe	TTC	33	79	55
	AAT	78	42	46		TTT	67	21	45
Asp	GAC	13	64	53	Pro	CCA	60	33	28
	GAT	87	36	47		CCC	10	27	32
Cys	TGC	10	60	55		CCG	2	6	11
	TGT	90	40	45		CCT	29	35	28
Gln	CAA	94	36	26	Ser	AGC	5	27	24
	CAG	6	64	74		AGT	24	7	15
Glu	GAA	94	38	42		TCA	42	7	15
	GAG	6	62	58		TCC	6	24	22
Gly	GGA	48	25	25		TCG	2	2	6
	GGC	7	43	34		TCT	21	33	18
	GGG	0	15	24	Thr	ACA	57	28	28
	GGT	46	18	17		ACC	6	41	36
His	CAC	26	57	59		ACG	2	8	12
	CAT	74	43	41		ACT	35	22	24
Ile	ATA	41	12	15	Trp	TGG	0	0	100
	ATC	6	55	49		TAC	19	54	56
	ATT	53	34	36	Tyr	TAT	81	46	44
Val						GTA	45	6	12
					Val	GTC	5	33	24
						GTG	4	48	46
						GTT	46	13	18

The synthetic MSP-1 sequence was adjusted to the human codon frequencies using a random number generator. Multiple additional adjustments, e.g. for generating unique cleavage sites, eliminating splice donor and acceptor signals, etc., were made thereafter. The codon frequencies shown for the synthetic gene represent the final sequence synthesized which, when compared with the native gene, contains base pair changes at 1317 positions.

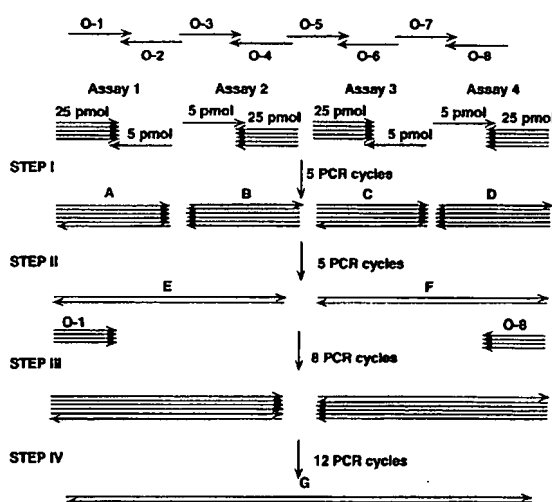
Searching for hidden recognition sequences for restriction endonucleases permitted us, again by incorporating single base pair exchanges, to position unique cleavage sites at or near the major processing sites where the MSP-1 precursor of FCB-1 is proteolytically cleaved during schizogony (Fig. 1A). Thus, the endonucleases *SphI*, *BstEII* and *Eco47III* cleave the synthetic MSP-1 coding sequence within one, four and one amino acid, respectively, of the processing sites which separate p19 from p29, p29 from p38 and p38 from p30 (Fig. 1A). Moreover, the *XmnI* site at position 2025 is within 28 amino acids of the putative processing site that separates p83 from p30 (34). Finally, unique cleavage sites were placed at either end of the full-size gene: a *MluI* site at the 5'-end and a *NotI* as well as a *Clal* site at the 3'-end of the gene (Fig. 1B).

Because MSP-1<sup>S2</sup> (Fig. 1B and C) lacks the signal peptide and the anchoring signal, it should be synthesized and remain in the cytoplasm. We, therefore, prepared two further genes with modifications to the 5'- and 3'-ends. In one (*msp-1<sup>S3</sup>*), the original signal peptide at the 5'-end and the GPI anchor signal at the 3'-end flank the coding sequence for the mature protein. In the other (*msp-1<sup>S4</sup>*), the signal sequence, but not the anchor sequence, which should permit secretion of the protein but not membrane retention, flank the coding sequence for the mature protein. While modifications at the 5'-end of the gene are facilitated by a unique *HindIII* site at position 116, those at the 3'-end are best achieved by synthesizing variants of p19 which can be introduced via the unique *SphI* site.

The polynucleotide encoding MSP-1, complete with a signal peptide and GPI anchor signal, comprises 4917 bp whereas the



**Figure 1.** Schematic outline of the primary structure of MSP-1 of *P. falciparum* (FCB-1) and strategy for the synthesis of its coding sequence. (A) The protein comprises 1639 amino acids including the signal peptide (SP) and the signal for GPI anchoring (GA). Conserved regions are depicted in white, dimorphic regions in grey. The two blocks showing the highest variability are hatched. Upper arrows delineate the major processing products corresponding to p83–p19 as well as, according to the nomenclature of Stafford *et al.* (34), SP and GA. The lower arrows indicate unique cleavage sites in the synthetic gene. They permit subdivision of the gene into segments encoding the individual processing products. The processing site between p83 and p30 has not been defined experimentally. The sizes of the sequences in bp encoding the various processing products of MSP-1 are depicted below. They allow calculation of molecular weights which for various reasons can significantly deviate from those derived previously by electrophoretic mobility of the respective proteins (2). (B) Flow chart of the gene synthesis. Five DNA fragments (83–19) were synthesized which encode the major processing products. They overlap their adjacent fragment by an average of 18 bp, which allows the fusion of neighbouring fragments via common unique endonuclease cleavage sites as indicated. Fragment 83 encoding the signal peptide contains at its 5'-end an *Mlu*I cleavage site (>). The 5'-end of all other fragments contains a *Bam*HI site (<). The 3'-end of p19 not encoding GA is followed by a *Not*I (O) and a *Cla*I (●) site whereas the version of p19 encoding GA contains only a *Cla*I site. The fragments were fused stepwise as indicated to yield *msp-1<sup>S1</sup>*. The GA signal of the parasite was introduced by an appropriately modified fragment 19 to yield *msp-1<sup>S</sup>*, whereas the SP sequence was eliminated from *msp-1<sup>S1</sup>* by inserting the appropriate oligonucleotide between the *Mlu*I and a unique *Hind*III site at position 116. The resulting *msp-1<sup>S2</sup>* gene can be inserted into expression vectors via its unique *Bam*HI and *Cla*I (or *Not*I) cleavage sites. (C) N- and C-terminal sequences of MSP-1<sup>S</sup>, MSP-1<sup>S1</sup> and MSP-1<sup>S2</sup> at the nucleotide and amino acid levels. The numbering of the amino acid positions is according to Heidrich *et al.* (24).



**Figure 2.** Flow chart for the synthesis of a polynucleotide of 600–1100 bp in length. Eight synthetic oligonucleotides (O1–O8) which overlap their respective neighbouring sequences by an average of 18 nt were mixed pairwise in four assays at the stoichiometry indicated. After five amplification cycles, products A–D are obtained. Fragment A was combined with B and C with D, producing E and F, respectively, after another five additional amplification cycles. Asymmetry in DNA strand composition was reintroduced by amplifying E in the presence of O1 and F in the presence of O8. The final product G was prepared by combining the asymmetric mixtures of E and F and amplifying for 12 cycles. The product was purified from agarose gels. For fragments where 10 or 12 oligonucleotides encoding p83, purified amplification products corresponding to G and D or F were subjected to amplification steps III and IV to yield the final products such as p83 DNA fragments I and II.

total sequence synthesized including two stop codons and flanking restriction cleavage sites comprised 4940 bp.

#### Synthesis and cloning of polynucleotides encoding MSP-1 or portions thereof

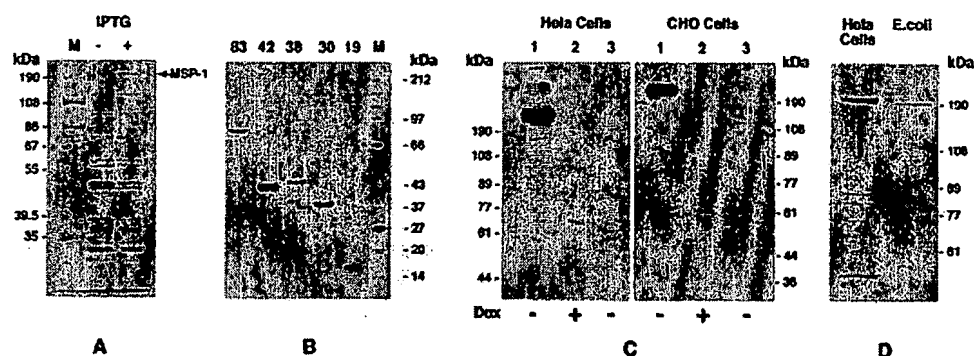
The full-length *msp-1* gene sequence, designed above, was subdivided into five overlapping fragments each corresponding to one of the major processing products of MSP-1, namely p83, p30, p38, p29 and p19 (Fig. 1B). They were synthesized using the PCR-based procedure outlined in Figure 2 that allows efficient production of double-stranded DNA fragments of up to 1200 bp long. With the exception of DNA encoding p83, each fragment was prepared from four pairs of overlapping synthetic oligonucleotides that covered the fragment as part of the upper or lower strand. To facilitate later assembly of adjacent fragments, the terminal O1 oligonucleotides, which encoded the fragment N-termini, also contained the unique 3' cleavage sites from the upstream adjacent fragment and the terminal O8 oligonucleotides, encoding the fragment C-termini, contained the unique 5' cleavage sites from the downstream adjacent fragments. Moreover, every O1 oligonucleotide contained a *Bam*HI site upstream of its unique 5' cleavage site. Similarly, every O8 oligonucleotide contained two tandem stop codons and a *Cla*I site or a *Not*I and a *Cla*I site downstream of the unique cleavage site (Fig. 1B and C). These features allow the cloning of each individual fragment encoding a processing product of MSP-1 via *Bam*HI/*Cla*I

cleavage. An exception to this is the DNA fragment encoding p83, which contained the N-terminal signal peptide. In this case, the 5'-end of O1 contains a *Mlu*I site immediately upstream of the start codon of the full-size gene. This *Mlu*I site allows us to transfer the assembled MSP-1-encoding DNA into any vector via the unique *Mlu*I/*Cla*I or *Mlu*I/*Not*I sites. The synthesis of the DNA fragment encoding p83 and the N-terminal signal sequence required another modification of the scheme due to the size of 2025 bp of the oligonucleotide. The more N-terminal fragment I was synthesized from a total of 12 oligonucleotides of 106–126 bp. Eight of these produced the 5' portion (778 bp) of fragment I according to Figure 2. The 445 bp 3' portion was synthesized from four oligonucleotides according to steps I and II in Figure 2. The complete fragment I was then generated by joining these two fragments by PCR as outlined for steps III and IV in Figure 2. The p83 DNA fragment II (954 bp) was generated from 10 oligonucleotides in an analogous procedure. Both p83 DNA fragments I and II were, after sequence verification, combined via a unique *Sca*I site at position 1145.

The full-size gene encoding the entire MSP-1 was obtained by successively joining the various DNA fragments corresponding to the processing products as outlined in Figure 1B. The polynucleotide encoding the entire MSP-1<sup>S1</sup> was generated by combining the fragment encoding p83 and p30 with the fusion products encoding p38 and p42 via the *Eco*47III cleavage site. Sequences encoding signals for GPI anchoring were attached by inserting properly modified polynucleotides via the *Sph*I/*Cla*I cleavage site whereas modifications of the N-terminal signal peptide were generated by exchanging sequences upstream of the unique *Hind*III site (data not shown). The synthetic gene which includes the coding sequence for the authentic signal peptide and for the GPI-mediated anchoring signal of the parasite is designated *msp-1*<sup>S</sup> (Fig. 1C). For expression studies in *E.coli*, the sequence encoding the signal peptide was removed, yielding *msp-1*<sup>S2</sup>. This was achieved by replacing the *Mlu*I–*Hind*III fragment by an appropriately synthesized *Bam*HI–*Hind*III oligonucleotide as outlined in Figure 1B and C. The resulting gene starts with an ATG followed by the codon for amino acid 21 (Fig. 1C). All the sequences described herein were submitted to the EMBL database under accession no. AJ131294.

#### Controlled expression of *msp-1*<sup>S2</sup> DNA in *E.coli* and isolation of the products

The *msp-1*<sup>S2</sup> as well as portions thereof encoding p83, p30, p38, p42 and p19 were placed under the control of an IPTG-inducible promoter in a vector that fuses six histidines to the N-terminus of the expression products. Upon induction, all six proteins were produced in *E.coli* and cell extracts were subjected to Ni<sup>2+</sup> chelate chromatography under denaturing conditions. Renaturation of the adsorbed material via a urea gradient and subsequent elution with imidazole hydrochloride yielded the six proteins in highly purified and soluble form. The induction of MSP-1<sup>S2</sup> synthesis in *E.coli* and electrophoretic characterization of purified p19–p83 is shown in Figure 3A. The isolation procedure via an N-terminal histidine tag was chosen since several fragments do not contain the epitope recognized by the 5.2 mAb antibody most suitable for immunoaffinity chromatography. Moreover, high expression levels may lead to inclusion of bodies from which the respective proteins can be readily recovered under denaturing conditions. Adsorption of such proteins to Ni<sup>2+</sup> chelate supports facilitates their renaturation.



**Figure 3.** Expression of *msp-1<sup>S2</sup>* and *msp-1* fragments in *E. coli* and in mammalian cells. (A) The *msp-1<sup>S2</sup>* sequence was inserted into expression vector pDSS6 where it is under control of an IPTG inducible promoter. Upon transfer in *E. coli* strain SG13009, the synthesis of MSP-1<sup>S2</sup> can be induced (IPTG+). The arrow identifies an induced product with a molecular weight of ~190 kDa. M denotes a molecular weight marker. (B) Electrophoretic analysis of IPTG-induced MSP-1 fragments produced in *E. coli* and purified by Ni<sup>2+</sup> chelate chromatography. The various fragments are indicated above the lane. The designation of the protein fragments (34) is not identical with their molecular mass calculated from the respective sequences (Fig. 1). In addition, some (p38, p30 and p19) show a migration behaviour that does not allow a strict correlation with the molecular weight standard. M denotes a broad range molecular weight marker (New England Biolabs). Coomassie stained 4–12% PAA gradient gel. (C) Expression of MSP-1 in HeLa and CHO cells. HtTA-93/9 and CtTA27/29 cells were grown to 40% confluency in the presence of Dox (100 ng/ml) before the antibiotic was removed. After 24 h, cell extracts were prepared and analysed by western blot using mAb 5.2. Lanes 1 and 2 show extracts from induced (–Dox) and uninduced (+Dox) cultures, lane 3 shows extracts from HtTA-1 and CtTA cells. (D) Electrophoresis of purified MSP-1 from HeLa cells and from *E. coli*. The left lane shows a MSP-1<sup>S1</sup> preparation obtained from cultures of HtTA-93/9 cells after immunoaffinity purification using the 5.2 mAb and the right lane shows full-size material from *E. coli* obtained by Ni<sup>2+</sup> chelate chromatography. Coomassie stained PAA gels.

### Production of MSP-1<sup>S1</sup> in mammalian cells

It may be difficult to obtain in *E. coli* the properly folded form of a protein like MSP-1 that is normally secreted and membrane anchored. We therefore also studied the expression of the synthetic genes encoding MSP-1<sup>S1</sup> and MSP-1<sup>S2</sup> in HeLa and CHO cells. Since preliminary results suggested that MSP-1 may interfere with the cellular metabolism (data not shown), its synthesis was controlled via the tetracycline regulatory system (31). Thus, the coding sequences were placed under the control of a bidirectional promoter (30) that is responsive to the tetracycline-controlled transcriptional transactivator (tTA). In these constructs, the expression of the *msp-1* gene is co-regulated with the luciferase reporter gene which is used as a convenient screening tool for identifying stably transformed cell lines. HeLa and CHO cell lines for the controlled expression of the *msp-1<sup>S1</sup>* gene were generated by transfecting HtTA-1 and CtTA-1 cells, which constitutively produce tTA (31), with the appropriate plasmids. Clones that showed good tetracycline-dependent regulation of luciferase were selected and examined for MSP-1<sup>S1</sup> synthesis. Several HeLa and CHO cell lines such as HtTA-93/9 and CtTA-27/29 were established. They exhibited high regulation factors for luciferase (up to 10<sup>4</sup>-fold) and also co-regulate well the synthesis of MSP-1<sup>S1</sup> (Fig. 3B). Interestingly, although the *msp-1* gene encodes the genuine signal peptide, no protein could be recovered from the culture supernatant of both HtTA-93/9 and CtTA-27/29 cells, suggesting that the protein is not liberated under these conditions. HtTA-93/9 cells grown at preparative scale allowed us to isolate full-length MSP-1 from cell extracts via immunoaffinity chromatography (Fig. 3D) on columns prepared with mAb 5.2.

### Interaction of heterologously produced MSP-1 with monoclonal antibodies directed against the native protein

To gain a first insight into conformational properties of MSP-1 as isolated from *E. coli* and from mammalian cells, purified antigen was reacted with a panel of MSP-1-specific monoclonal antibodies. Of these antibodies, five are specific for the K1 prototype as represented by the FCB-1 sequence, six are known to recognize epitopes within the conserved parts of the molecule and two are specific for MAD20 sequences. Nine of the antibodies react with conformational rather than with sequential epitopes (35,36). When preparations of MSP-1<sup>S1</sup> from CHO cells and MSP-1<sup>S2</sup> from *E. coli* were probed with the various antibodies in western blots, a rather striking result was obtained. MSP-1<sup>S2</sup> isolated from *E. coli* interacted with all 11 antibodies that are specific for the K1 prototype. No interaction was seen with antibody 9.7 specific for the MAD20 prototype or with antibody 12.1, which recognizes an oligomeric sequence of block IV which is not present in MSP-1 of the FCB-1 strain (Fig. 4 and Table 2). In contrast, MSP-1<sup>S1</sup> isolated from CHO cells while interacting with most K1-specific antibodies was not recognized by three of the mAbs that bind to MSP-1<sup>S2</sup> isolated from *E. coli*. Again, as expected, there was no interaction between CHO-derived MSP-1<sup>S1</sup> and monoclonal antibodies 9.7 and 12.1 (Fig. 4 and Table 2).

### DISCUSSION

DNA of *Pfalciparum* has an extraordinarily high AT content which can exceed 90% in intragenic regions and may reach 75% in coding sequences. The reasons for the strong preference of AT over GC, most clearly revealed at the wobble position of codons (Table 1), are not understood. A consequence of the high AT

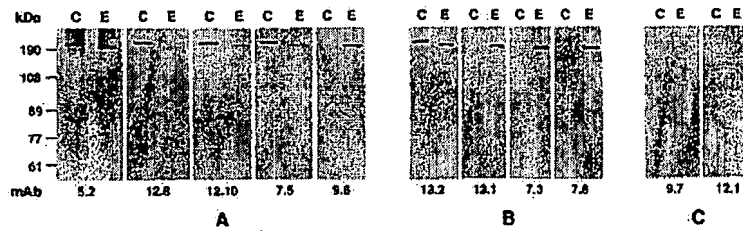


Figure 4. Interaction of MSP-1 from *E. coli* and CHO cells with monoclonal antibodies. MSP-1<sup>S2</sup> and MSP-1<sup>S1</sup> isolated from *E. coli* and CHTA-27/29 cells, respectively, were subjected to electrophoresis under non-reducing conditions and analysed by western blot using the monoclonal antibodies indicated. MSP-1<sup>S1</sup> from CHO cells (C) migrates slower than MSP-1<sup>S2</sup> from *E. coli* (E) due to glycosylation in the eukaryotic expression system. (A) Antibodies directed towards conserved portions of MSP-1. Interactions with antibodies specific for dimorphic regions are shown in (B) (K1) and (C) (MAD20).

Table 2. MSP-1 purified from *E. coli* and CHO cells, respectively, was subjected to gel electrophoresis under non-reducing conditions

Specificity	MSP-1 specific antibodies				IFA		
	mAb No.	Epitope Type	Region of interaction	Ref.	<i>P.f.</i> FCB-1	<i>E. coli</i>	CHO
Conserved	5.2	conformational	p19	32	++++	+	+
	12.8		p42	35,36	++	+	+
	12.10	conformational	p19	35	++++	+	+
	7.5		p19	35	++++	+	+
	9.8			35,36	++++	+	-
	2.2		p19	35	++++	+	+
K1	13.2		p83	35,36	++++	+	+
	13.1		p42	35,36	++++	+	-
	7.3	conformational		35	++++	+	-
	7.6	conformational		35	++++	+	+
	6.1		p42	35,36	++++	+	ND
MAD20	9.7	conformational		36	-	-	-
	12.1		p83 (block IV)	36	-	-	-

Immunoblots with a panel of well-characterized monoclonal antibodies (32,35,36) were prepared and visually evaluated (Fig. 4). Indirect immunofluorescence assays (IFA) were performed as described in Harlow and Lane (33). The number of + reflects the intensity of immunofluorescence. ND, experiment not done due to limiting amounts of mAb 6.1. The western blots for mAb 2.2 and 6.1 are not shown in Figure 4 since limiting amounts of the antibodies did not allow us to carry out this experiment more than once.

content of *Pfalciparum* DNA is the failure of cloning and stably maintaining large genes in *E. coli* rendering in-depth studies of the respective gene functions most difficult. In some cases, the biased codon composition was even believed to hamper the expression of *Pfalciparum* genes in heterologous systems (37,38).

The synthesis of a 4917 bp long polynucleotide encoding the 190 kDa MSP-1 of the FCB-1 strain in a codon composition that reduces the AT content to 55% has opened up new possibilities for the study of this intriguing protein since the synthetic gene can now be stably cloned and expressed in its entirety in *E. coli* as well as in a variety of other heterologous systems. Several parameters were reconciled in the design of the synthetic *msh-1* gene. Thus, it appeared sensible to place unique endonuclease cleavage sites at or near positions where the protein is processed at the surface

of the mature schizont which permits separate cloning and expression of the portions of the *msh-1* gene that encode the different processing products (Fig. 1). Although, with the exception of the GPI anchor at the C-terminus, native MSP-1 of *Pfalciparum* appears not to be glycosylated (39) we have conserved potential glycosylation sites as any change in the amino acid composition may destroy epitopes important in the host-parasite interaction. Moreover, choosing proper hosts for MSP-1 synthesis can prevent the modification of such sites and, finally, it will not be difficult to eliminate glycosylation sites at a later stage should it become desirable.

The unique *HindIII* site near the 5'-end and the *SphI* site near the 3'-end of the gene allow for switching signal peptides or membrane anchoring signals. Indeed, besides the sequence



modifications shown here (MSP-1<sup>S</sup> to MSP-1<sup>S2</sup>, Fig. 1), we have fused sequences encoding several other specific signal peptides or anchoring signals with the gene (data not shown). Further parameters that were considered are discussed in Results and Materials and Methods.

For synthesizing an oligonucleotide of the size of the *msh-1* gene, we have examined several strategies. The approach described here is based on an asymmetric amplification process that starts out with eight overlapping oligonucleotides and leads to an end product of 600–800 bp, without the requirement of isolating intermediates. Following this strategy, it is essential to limit the size of the starting oligonucleotides to <120, optimally to 70–90 nt, since the error rate of the PCR products increases for longer oligonucleotides, most likely due to incomplete deprotection or to modifications of nucleotides during chemical synthesis. The complete *msh-1<sup>S2</sup>* gene and all the synthetic intermediates obtained by this procedure were stably cloned in *E.coli* confirming that the parameter responsible for the instability of large *P.falciparum* genes in *E.coli* is the high AT content.

First expression studies with the synthetic sequences in *E.coli* revealed that MSP-1<sup>S2</sup> was readily produced as an intracellular protein. Moreover, N-terminal fusion of a histidine tag allows its rapid isolation in soluble form via affinity chromatography. Encouragingly, the examination of such MSP-1<sup>S2</sup> preparations with a panel of monoclonal antibodies, of which several are considered to recognize conformational epitopes, indicates that at least some portions of the protein are properly folded under the conditions used. Production and purification of MSP-1 fragments corresponding to the various processing products of the native protein are even more efficient and all the fragments are obtained as soluble proteins, a prerequisite for structural and functional studies.

Despite the promising results in *E.coli*, proper folding of a complex protein like MSP-1 that is transported to the surface of the parasite may be more readily achieved in a eukaryotic system under conditions of secretion and possibly membrane anchoring. We have, therefore, begun to study the expression of the *msh-1<sup>S1</sup>* gene in mammalian cells. After initial studies suggested that synthesis of MSP-1 may negatively affect the growth of HeLa cells, we placed the gene under tetracycline control and generated stable HeLa and CHO cell lines where MSP-1<sup>S1</sup> synthesis is stringently controlled and can be induced over several orders of magnitude. Full-size protein is recovered from cell extracts upon induction. Since the *msh-1<sup>S1</sup>* gene encodes the genuine signal peptide of the parasite, which is quite similar to other eukaryotic signal sequences, one might anticipate the secretion of the protein into the culture supernatant. We failed, however, to detect any secreted material (data not shown) and are presently analysing in which compartment of the cell the protein accumulates. The isolated protein migrates distinctly slower in an electric field than the protein produced in *E.coli* and thus it can be assumed that it enters the endoplasmic reticulum and the Golgi pathway where it is glycosylated. Interestingly, full-size MSP-1 can be isolated from cell extracts by affinity chromatography with mAb 5.2, i.e. by an antibody that recognizes a conformational epitope near the C-terminus. This indicates again that at least this rather critical domain of the protein may be in a conformation identical to the native one. This conclusion is supported by the reactivity of the antigen with mAbs 12.10, 7.5 and 2.2, all of which have been mapped to conformation-dependent epitopes within p19 of MSP-1 (Table 2).

The availability of a *msh-1* gene that can be transferred, stably maintained and expressed in various biological systems will advance the elucidation of its role in the parasite's life cycle. This will include structural analysis of the intact protein as well as of its processing products. Of particular interest will be the analysis of the interaction of MSP-1 and its processing products with erythrocytes.

Several approaches are under way in our laboratory. For example, we have successfully placed full-size MSP-1 as well as portions thereof onto the surface of HeLa cells (P.Burghaus, manuscript in preparation) and *Toxoplasma gondii* where they are anchored by a GPI moiety (I.Türbachova *et al.*, manuscript in preparation). Interaction with our collection of mAbs suggests that the surface-exposed proteins have assumed the natural conformation. These systems are opening up a variety of experimental strategies aimed at the analysis of MSP-1 function. This will include not only the interaction of MSP-1 or any portion thereof with the surface of erythrocytes but also questions concerning the maturation of the MSP-1 precursor. Thus, heterologously produced MSP-1 or portions thereof may constitute useful substrates for proteases involved in this process (41). Finally, it will be possible to study the interaction between purified MSP-1 fragments representing the natural processing products which may allow reassembly of the MSP-1 complex *in vitro*. Together, such studies may lead to new insights into the early phases of erythrocyte invasion and reveal new targets for interfering with *P.falciparum* infection at the blood stage.

The synthetic *msh-1* gene will also facilitate the rigorous examination of the protective potential of MSP-1 or any portion thereof when used as an experimental vaccine. Most of the recent MSP-1-based vaccination trials focused on the C-terminal portion of the protein, in part for technical reasons. While such studies have clearly identified this region of the molecule as promising, the analysis can now be extended throughout the entire molecule as there is little reason to exclude any portion of this surface protein from such examination, particularly considering the contribution of cellular responses towards malaria immunity where MSP-1 could play a role (40). The FCB-1-derived amino acid sequence encoded in our synthetic gene is highly homologous to MSP-1 of the Colombian FVO strain which is well adapted to *Aotus* monkeys. Novel experimental vaccines that include the entire protein or any portion thereof can now be tested in this animal model and first monkey trials with vaccines based on heterologously expressed proteins as well as a recombinant attenuated *T.gondii* are under way. The sequence chosen here belongs to the K1 prototype. It diverges maximally from the MSP-1 sequence of the 3D7 strain, a representative of the MAD20 prototype. The synthesis of the gene encoding the 3D7 MSP-1 is presently being completed in our laboratory. Together, the two genes will allow comparative structural and immunologic studies. In particular, they will permit vaccination studies with heterologous challenge infections in *Aotus* monkeys. Moreover, the availability of unlimited amounts of MSP-1 proteins representing the processing products or any other portion of the K1- and the MAD20-derived MSP-1 will make a detailed analysis of the humoral response in populations where *P.falciparum* is endemic feasible. Such analyses may lead to more reliable correlations between patterns of humoral response and susceptibility towards infection and disease as was suggested in earlier studies (10,11), possibly allowing for the development of diagnostic tools with

predictive value. Indeed, analysing sera from recent immunization trials with *Aotus* monkeys using the various MSP-1 fragments have revealed clear correlations between antibodies directed towards certain areas of the protein and protection (R. Tolle *et al.*, manuscript in preparation).

Finally, since the well-characterized 3D7 strain is being used in human trials for challenge infections, immunization of humans followed by homologous or heterologous challenge appears feasible in the near future. Together, such studies should increase our understanding of the role of MSP-1 in the parasite's life cycle, the basis of its dimorphic nature and its potential as a component in a subunit vaccine.

## ACKNOWLEDGEMENTS

We would like to thank J. McBride for supplying monoclonal antibodies and S. Freundlieb, U. Baron and C. Schirra for help in expression studies. Helpful suggestions by M. Nassal and support by R. Sprengel in automated DNA sequencing are gratefully acknowledged. We are greatly indebted to J. Lyon and H. del Portillo for suggestions and critical reading of the manuscript. This work was supported by grants of the Federal Ministry of Research and Technology (BMFT/BMBF AZ KA 9301) and by the Fonds der Chemischen Industrie Deutschlands.

## REFERENCES

- Holder, A.A. and Freeman, R.R. (1982) *J. Exp. Med.*, **156**, 1528–1538.
- Holder, A.A. (1988) *Prog. Allergy*, **41**, 72–97.
- Tanabe, K., Mackay, M., Goman, M. and Scaife, J.G. (1987) *J. Mol. Biol.*, **195**, 273–287.
- Cooper, J.A. (1993) *Parasitol. Today*, **9**, 50–54.
- Miller, L.H., Roberts, T., Shahabuddin, M. and McCutchan, T.F. (1993) *Mol. Biochem. Parasitol.*, **59**, 1–14.
- Holder, A.A. and Freeman, R.R. (1981) *Nature*, **294**, 361–364.
- Majarian, W.R., Daly, T.M., Weidanz, W.P. and Long, C.A. (1984) *J. Immunol.*, **132**, 3131–3137.
- Lew, A.M., Langford, C.J., Anders, R.F., Kemp, D.J., Saul, A., Fardoulis, C., Geysen, M. and Sheppard, M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 3768–3772.
- Ling, J.T., Ogun, S.A., Momin, P., Richards, R.L., Garcon, N., Cohen, J., Ballou, W.R. and Holder, A.A. (1997) *Vaccine*, **15**, 1562–1567.
- Tolle, R., Früh, K., Doumbo, O., Koita, O., N'Diaye, M., Fischer, A., Dietz, K. and Bujard, H. (1993) *Infect. Immun.*, **61**, 40–47.
- Egan, A.F., Chappell, J.A., Burghaus, P.A., Morris, J.S., McBride, J.S., Holder, A.A., Kaslow, D.C. and Riley, E.M. (1995) *Infect. Immun.*, **63**, 456–466.
- Perrin, L.H., Merkli, B., Loche, M., Chizzolini, C., Smart, J. and Richle, R. (1984) *J. Exp. Med.*, **160**, 441–451.
- Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., Mackay, M. and Scaife, J.G. (1984) *Nature*, **311**, 379–382.
- Siddiqui, W.A., Tam, L.Q., Kramer, K.J., Hui, G.S.N., Case, S.E., Yamaga, K.M., Chang, S.P., Chan, E.B.T. and Kan, S.-C. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 3014–3018.
- Eitlinger, H.M., Caspers, P., Matile, H., Schoenfeld, H.-J., Stueber, D. and Takacs, B. (1991) *Infect. Immun.*, **59**, 3498–3503.
- Holder, A.A., Freeman, R.R. and Nicholls, S.C. (1988) *Parasite Immunol.*, **10**, 607–617.
- Herrera, S., Herrera, M.A., Perlaza, B.L., Burki, Y., Caspers, P., Döbeli, H., Rotmann, D. and Certa, U. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4017–4021.
- Herrera, M.A., Rosero, F., Herrera, S., Caspers, P., Rotmann, D., Sinigaglia, F. and Certa, U. (1992) *Infect. Immun.*, **60**, 154–158.
- Patarroyo, M.E., Romero, P., Torres, M.L., Clavijo, P., Moreno, A., Martínez, A., Rodríguez, R., Guzmán, F. and Cabezas, E. (1987) *Nature*, **328**, 629–632.
- Chang, S.P., Case, S.E., Gosnell, W.L., Hashimoto, A., Kramer, K.J., Tam, L.Q., Hashiro, C.Q., Nikaido, C.M., Gibson, H.L., Lee-Ng, C.T., Barr, P.J., Yokota, B.T. and Hui, G.S.N. (1996) *Infect. Immun.*, **64**, 253–261.
- Burghaus, P.A., Wellde, B.T., Hall, T., Richards, R.L., Egan, A.F., Riley, E.M., Ballou, W.R. and Holder, A.A. (1996) *Infect. Immun.*, **64**, 3614–3619.
- Pye, D., Edwards, S.J., Anders, R.F., O'Brien, C.M., Franchina, P., Corcoran, L.N., Monger, G., Peterson, M.G., Vandenberg, K.L., Smythe, J.A., Westley, S.R., Coppel, R.L., Webster, T.L., Kemp, D.J., Hampson, A.W. and Langford, C.J. (1991) *Infect. Immun.*, **59**, 2403–2411.
- Sandhu, J.S. and Kennedy, J.F. (1994) *Vaccine*, **12**, 56–64.
- Heidrich, H.-G., Miettinen-Baumann, A., Eckerskorn, C. and Lottspeich, F. (1989) *Mol. Biochem. Parasitol.*, **34**, 147–154.
- Pan, W., Tolle, R. and Bujard, H. (1995) *Mol. Biochem. Parasitol.*, **73**, 241–244.
- Gribskov, M., Devereux, J. and Burgess, R.R. (1984) *Nucleic Acids Res.*, **12**, 539–549.
- Devereux, J., Haeberli, P., Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
- Sandhu, G.S., Alef, R.A. and Kline, B.C. (1992) *Biotechniques*, **12**, 14–16.
- Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. and Stüber, D. (1988) *BioTechnology*, **6**, 1321–1325.
- Baron, U., Freundlieb, S., Gossen, M. and Bujard, H. (1996) *Nucleic Acids Res.*, **23**, 3605–3606.
- Gossen, M. and Bujard, H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5547–5551.
- Siddiqui, W.A.L., Tam, S.C., Kan, K.J., Kramer, S.E., Case, K.L., Palmer, K.M., Yamaga, G. and Hui, G. (1986) *Infect. Immun.*, **52**, 314–318.
- Harlow, E. and Lane, D. (1988) *Antibodies—A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Ch. 12 and 13, pp. 505–523.
- Stafford, W.H.L., Blackman, M.J., Harris, A., Shafira, S., Grainger, M. and Holder, A.A. (1994) *Mol. Biochem. Parasitol.*, **66**, 157–160.
- McBride, J.S. and Heidrich, H.G. (1987) *Mol. Biochem. Parasitol.*, **23**, 71–84.
- Conway, D.J., Rosario, V., Oduola, A.M., Salako, L.A., Greenwood, B.M. and McBride, J.S. (1991) *Exp. Parasitol.*, **73**, 469–480.
- Sugiyama, T., Suzue, K., Okamoto, M., Inselburg, J., Tai, K. and Horii, T. (1996) *Vaccine*, **14**, 1069–1076.
- Prapunwattana, P., Sirawaraporn, W., Yuthavong, Y. and Santi, D.V. (1996) *Mol. Biochem. Parasitol.*, **83**, 93–106.
- Berhe, S. (1998) PhD thesis, Universität Marburg.
- Krzych, U., Lyon, J.A., Jareed, T., Schneider, I., Hollingdale, M.R., Gordon, D.M. and Ballou, W.R. (1995) *J. Immunol.*, **155**, 4072–4077.
- Blackman, M.J., Heidrich, H.G., Donachie, S., McBride, J.S. and Holder, A.A. (1990) *J. Exp. Med.*, **172**, 379–382.

# The BIOCHEMICAL JOURNAL

The Biochemical Society, London © 1991

**Volume 279**

**1991**

## Editorial Board

### Chairman

A. J. Turner (*Leeds*)

### Deputy Chairmen

W. H. Evans (*London*), *reviews editor*

H. G. Nimmo (*Glasgow*)

K. Snell (*Guildford*)

G. H. Elder (*Cardiff*)

S. J. Higgins (*Leeds*)

B. Burchell (*Dundee*)

### Managing Editor

A. S. Beedle

G. Allen (*Beckenham*)

B. H. Anderton (*London*)

D. J. Anstee (*Bristol*)

G. Banting (*Bristol*)

T. Berg (*Oslo*)

S. J. W. Busby (*Birmingham*)

R. J. Cherry (*Colchester*)

G. M. W. Cook (*Cambridge*)

A. P. Dawson (*Norwich*)

H. R. de Jonge (*Rotterdam*)

A. H. Drummond (*Oxford*)

J. Feeney (*London*)

D. A. Fell (*Oxford*)

M. A. J. Ferguson (*Dundee*)

S. J. Ferguson (*Oxford*)

J. B. C. Findlay (*Leeds*)

G. F. Gibbons (*Oxford*)

J. P. Goddard (*Glasgow*)

L. Hall (*Bristol*)

B. Halliwell (*Sacramento*)

D. G. Hardie (*Dundee*)

T. E. Hardingham (*London*)

D. Häussinger (*Freiburg*)

H. Holmsen (*Bergen*)

R. F. Irvine (*Cambridge*)

M. A. Kerr (*Dundee*)

U. Lindahl (*Uppsala*)

J. P. Luzio (*Cambridge*)

B. Mannervik (*Uppsala*)

S. B. Marston (*London*)

J. G. McCormack (*Edinburgh*)

J. D. McGivan (*Bristol*)

M. A. McNurlan (*Aberdeen*)

I. A. Nimmo (*Edinburgh*)

Y. Nishizuka (*Kobe*)

V. M. Pain (*Brighton*)

P. Parker (*London*)

A. E. Pegg (*Hershey*)

N. C. Price (*Stirling*)

C. G. Proud (*Bristol*)

H. Prydz (*Oslo*)

P. J. Reeds (*Houston*)

A. J. Rivett (*Leicester*)

S. P. Robins (*Aberdeen*)

S. B. Shears (*Research Triangle Park*)

K. Siddle (*Cambridge*)

H. Sies (*Düsseldorf*)

A. R. Slabas (*Durham*)

B. E. Smith (*Brighton*)

J. B. Smith (*Philadelphia*)

A. K. Soutar (*London*)

W. L. R. Stalmans (*Leuven*)

P. G. Strange (*Canterbury*)

L. M. G. van Golde (*Utrecht*)

M. J. O. Wakelam (*Glasgow*)

M. Wallis (*Brighton*)

M. Watford (*New Brunswick*)

C. W. Wharton (*Birmingham*)

V. A. Zammit (*Ayr*)

# Engineering a bioluminescent indicator for cyclic AMP-dependent protein kinase

Graciela B. SALA-NEWBY\* and Anthony K. CAMPBELL

Department of Medical Biochemistry, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, U.K.

cDNA coding for the luciferase in the firefly *Photinus pyralis* was amplified *in vitro* to generate cyclic AMP-dependent protein kinase phosphorylation sites. The DNA was transcribed and translated to generate light-emitting protein. A valine at position 217 was mutated to arginine to generate a site RRFS and the heptapeptide kemptide, the phosphorylation site of the porcine pyruvate kinase, was added at the *N*- or *C*-terminus of the luciferase. The proteins carrying phosphorylation sites were characterized for their specific activity, pI, effect of pH on the colour of the light emitted and effect of the catalytic subunit of protein kinase A in the presence of ATP. Only one of the recombinant proteins (RRFS) was significantly different from wild-type luciferase. The RRFS mutant had a lower specific activity, lower pH optimum, emitted greener light at low pH and when phosphorylated it decreased its activity by up to 80%. This latter effect was reversed by phosphatase. This recombinant protein is a good candidate to measure for the first time cyclic AMP-dependent phosphorylation in live cells.

## INTRODUCTION

A universal feature of eukaryotic cells is the ability of physiological agonists, such as hormones, growth factors and neurotransmitters, components of the body's defence system, non-host antigens and other pathogens and drugs, to interact with the plasma membrane and trigger molecular events within the cell. These agents initiate a molecular sequence that starts with the generation of an intracellular signal, such as  $\text{Ca}^{2+}$ , cyclic AMP, inositol trisphosphate or diacylglycerol, and ends with a physiological or pathological event in the cell (Campbell, 1983; Berridge & Irvine, 1989). These events include movement, secretion, transformation, division, defence and death. The timing and magnitude of the end response in each cell depends on the timing and location of both the intracellular signals and the covalent modifications they induce. A particular cell will only undergo an end response if the right sequence of molecular thresholds has occurred (Campbell, 1983, 1988, 1990).

Measurement and imaging of intracellular  $\text{Ca}^{2+}$  using fluorescent and bioluminescent indicators (Campbell, 1983; Cobbold & Rink, 1987) has established that one explanation for gross heterogeneity in individual cell responses is variation in the timing and the location of the intracellular  $\text{Ca}^{2+}$  signal. In neutrophils, for example, four subpopulations have been defined, including one group showing no response at all (Hallett *et al.*, 1990; Davies *et al.*, 1991). A major problem in elucidating the molecular basis of heterogeneity within a cell population is the lack of a method for measuring and manipulating covalent modification of proteins in live cells. The purpose of the work reported here was to engineer cyclic AMP-dependent protein

kinase phosphorylation sites into firefly luciferase, such that a change in colour and/or light intensity occurred after phosphorylation and dephosphorylation (Campbell, 1989).

Benzothiazole luciferases occur only in luminous beetles. They contain approx. 550 amino acids, and require ATP,  $\text{Mg}^{2+}$  and  $\text{O}_2$ , as well as a common luciferin, to generate light (Campbell, 1988). Just a few amino acid changes can cause the colour to shift from green to green-yellow, yellow or red (Wood *et al.*, 1989a,b). Recognition sites for protein kinase A (Cohen, 1988) have been added to  $\alpha$ -interferon to allow high-specific-activity labelling for binding studies (Li *et al.*, 1989). In a previous study the heptapeptide kemptide (LRRASLG) (Zetterqvist *et al.*, 1976; Kemp *et al.*, 1977) was chemically coupled to extracted luciferase from the firefly *Photinus pyralis*. *Photinus* luciferase (*Photinus*-luciferin: oxygen 4-oxidoreductase; EC 1.13.12.7) emits yellow light with a peak intensity at 565 nm. The coupled kemptide shifted the colour of the light emitted to the red and phosphorylation shifted it even further (Jenkins *et al.*, 1990). Here PCR was used followed by transcription-translation *in vitro* (Sala-Newby *et al.*, 1990a,b) to change an amino acid sequence VRFS (217–220) (de Wet *et al.*, 1987) to RRFS, or to add kemptide to the *N*- or *C*-terminus of the protein.

## MATERIALS AND METHODS

### Materials

Oligonucleotide primers were prepared using an Applied Biosystems 381A DNA synthesizer and purified as 'trityl-on' oligonucleotides (100, 101) or 'trityl off' (105, 107, 108, 113, 114, T7-K). Their sequences were as follows:

- 100: TCATCGCTGAATACAGTTAC (3' end antisense)
- 101: GGTAAATGGAAGACGCCAAAAAC (5' end sense)
- 105: CACCTAATACGACTCACTATAGGGAGAATGGAAGACGCCAAAAAC (5' end antisense including the T7 promoter)
- 107: AGAACTGCCTGCCGAGATACTCGCA (5' end sense, underlined bases generate R codon)
- 108: TGCGAGAATCTGCGGCAGGCAGTTCT (3' end antisense, underlined bases generate R codon)
- 113: CCTGTGCTAGCTTACCCAGGGAGGCCGCCGAGCAATTGGACTTTCC (3' end antisense with 21 bases coding for kemptide, a stop codon and a *SalI* restriction site)
- 114: GGCTCCCTGGGCGAAGACGCCAAAAAC (5' end sense, part of kemptide)
- T7-K: CACCTAATACGACTCACTATAGGGAGAATGCTGCGGCGGGCTCCCTGGGC (5' end sense, clamp, T7 promoter and part of the coding sequence for kemptide)

Abbreviations used: CL count, chemiluminescence count; KNt, luciferase with kemptide at *N*-terminus; KCt, luciferase with kemptide at *C*-terminus.

\* To whom correspondence and reprint requests should be addressed.

The coding sequence for firefly luciferase was isolated from a cDNA library (Sala-Newby *et al.*, 1990a,b). A 2400 bp *SalI* fragment was used as the target for amplification. Amplitaq DNA polymerase was from Perkin-Elmer Ltd., U.K.; T7 RNA polymerase was from Promega, nucleotides and Sephacryl S100 were from Pharmacia and Centricon 100 cartridges were from Amicon, U.K. [ $\gamma$ - $^{32}$ P]ATP (10–50 Ci/mmol), [ $\alpha$ - $^{32}$ P]UTP (3000 Ci/mmol), stabilized [ $^{35}$ S]methionine ( $\geq 1000$  Ci/mmol), RNAase inhibitor and rabbit reticulocyte lysate (N90) were purchased from Amersham International plc. Restriction enzymes, alkaline phosphatase (24 units/ $\mu$ l) and luciferin were from Boehringer Corp. Low-protein-binding ultrafiltration units, Ultrafree MC, were from Millipore Corp. Protein kinase A inhibitor (P-3294) and kemptide were from Sigma Chemical Co. All other AnalaR-grade reagents were from Sigma Chemical Co. and BDH Chemicals. The catalytic subunit of cyclic AMP-dependent protein kinase was generously given by Dr. K. J. Murray of Smith Kline Beecham, Welwyn, Herts., U.K.

### Preparation of DNA fragments

Addition of the T7 RNA polymerase promoter (TAATACGACTCACTATAGGGAGA) (Stoflet *et al.*, 1988) and the DNA sequence coding for kemptide (CTGCGGCGGGGGTCCCTGGGC), as well as mutation of two bases within the luciferase cDNA, were carried out using PCR (Saiki *et al.*, 1988), as previously described (Sala-Newby *et al.*, 1990b). Firefly (*Photinus pyralis*) cDNA (4 ng/ml) was amplified in a solution containing 10 mM-Tris/HCl (pH 8.3), 50 mM-KCl, 2 mM-MgCl<sub>2</sub>, 0.01% (w/v) gelatin, 0.2 mM each of the four deoxynucleoside triphosphates, 0.5  $\mu$ M of each oligonucleotide primer and 40 units of Amplitaq DNA polymerase/ml. The cycling reactions were carried out in a Perkin-Elmer thermal cycler. Each of the 25 cycles consisted of 1 min at 94 °C, 1 min at 55 °C, 2 min at 72 °C plus a 5 s extension on each cycle. Klenow fragment of *Escherichia coli* DNA polymerase (40 units/ml) was added after the completion of the 25 cycles, and incubated for 30 min at 37 °C.

The final product was extracted once with phenol/chloroform/3-methylbutan-1-ol (25:24:1, by vol.) and precipitated with 2 vol. of 7.5 M-ammonium acetate plus 2.5 vol. of ethanol. The DNA concentration was assessed visually from ethidium bromide-stained agarose gels by comparison with the bands of a standard DNA (Sambrook *et al.*, 1989).

Wild-type firefly luciferase DNA preceded by the T7 RNA polymerase promoter was prepared using oligonucleotide primers 100 and 105 (see under 'Materials'). The kemptide coding sequence was added to the 3' end of the firefly cDNA using primers 105 and 113. The incorporation of the kemptide coding sequence at the 5' end and the 2 bp change coding for the mutation V-217  $\rightarrow$  R were carried out in two stages (Higuchi, 1990). After the first amplification the primers were removed by filtration through Centricon 100 cartridges (Higuchi, 1990). The first stage of the introduction of kemptide at the N-terminus was carried out using oligonucleotide primers 114 and 100. The resulting DNA (4 ng/ml) was amplified for 25 cycles in the presence of oligonucleotide primers T7-K and 100 to produce the final product. The first stage of the preparation of the RRFS variant generated two fragments: 5' end fragment (643 bp) was generated by amplification with oligonucleotide primers 101 and 108 and the 3' end (1018 bp) with oligonucleotide primers 107 and 100. In the second stage the two fragments were mixed in equimolar amounts (2  $\mu$ g of total DNA/ml), denatured (1 min at 94 °C) and allowed to reanneal by decreasing the temperature at 5.7 °C/min to 37 °C in amplification mixture containing primers 105 and 100 followed by 1 min extension at 72 °C. Eight to twelve cycles of amplification under normal conditions followed.

### Formation of luciferase *in vitro*

PCR products (0.5–1.5  $\mu$ g/25  $\mu$ l of incubation mixture) were transcribed as previously described (Sala-Newby *et al.*, 1990b). The RNA capped using 0.5 mM-m<sup>7</sup>G(5')ppp(5')G and 0.1 mM-GTP was precipitated twice with 0.2 vol. of 7.5 M-ammonium acetate and 2.5 vol. of ethanol. The RNA (1–100 ng) in 2  $\mu$ l of 10 mM-Tris/HCl, 1 mM-EDTA (pH 7.4), 3  $\mu$ l of potassium acetate and magnesium acetate to optimize their concentration (90–110 and 1.6–2.0 mM final concns. respectively) and 5  $\mu$ l of rabbit reticulocyte lysate N90 were incubated for 1 h at 30 °C, and luciferase activity as chemiluminescent (CL) count was measured in a home-built luminometer (Campbell, 1988) for 10 s at room temperature in 229  $\mu$ l of 20 mM-Tris/acetate/0.3 mM-dithiothreitol/0.2 mM-EDTA/1 mg of BSA/ml/12 mM-magnesium acetate/1.5 mM-ATP, pH 7.75. The reaction was started by addition of luciferin to 0.2 mM final concentration (1 ng of extracted luciferase yields  $2.1 \times 10^8$  and  $2.9 \times 10^8$  CL counts/10 s in the presence of rabbit reticulocyte lysate and buffer respectively). The amount of protein synthesized was measured by including 15  $\mu$ Ci of [ $^{35}$ S]methionine/10  $\mu$ l of translation cocktail. Proteins were separated on SDS/9% (w/v) polyacrylamide gels under reducing conditions (Laemmli, 1970) followed by fluorography and exposure to preflashed X-ray film. The luciferase bands were excised from the gel, radioactivity was measured in a liquid-scintillation counter and the amount of protein was estimated taking into account the concentration of methionine (28  $\mu$ M) in the lysate.

### Phosphorylation of proteins

The proteins were synthesized in 100–250  $\mu$ l of rabbit reticulocyte mixture, precipitated in 64% saturated ammonium sulphate, resuspended in 100  $\mu$ l of 50 mM-Tris/Mes/1 mM-EDTA/0.3 mM-dithiothreitol (pH 7.8) for normal, and for protein with kemptide N-terminus (KNt) and kemptide at C-terminus (KCt) and pH 7.2 for RRFS, and subjected to gel filtration on a column (0.7 cm  $\times$  20 cm) packed with Sephacryl S100 and equilibrated in the corresponding buffer. Active fractions were pooled and concentrated by ultrafiltration. Protein was measured by the method of Lowry *et al.* (1951). BSA (fraction V) was used as standard.

The phosphorylation was carried out in a mixture containing 20 mM-Mes, 60 mM-sodium glycerol 2-phosphate, 30 mM-NaF, 10 mM-magnesium acetate, 1 mM-EDTA, 1 mg of BSA/ml, 1  $\mu$ g each of leupeptin and pepstatin/ml and 125  $\mu$ M-ATP, pH 6.8. Active fractions from the gel filtration (0.8–1.2 mg of protein/ml, of which approximately 0.1% was luciferase) were added together with 0.5  $\mu$ l of purified catalytic subunit of protein kinase A or catalytic subunit diluent (0.5 M-potassium phosphate/0.1% Tween-20, pH 6.8) per 40  $\mu$ l of mixture [the catalytic subunit can transfer 7 mmol of  $^{32}$ P/min per  $\mu$ l using 20  $\mu$ M-malantide as a substrate, as in Murray *et al.* (1990)]. The incubations were carried out at 30 °C for 10–20 min. Kemptide was also phosphorylated in the presence of rabbit reticulocyte that was gel-filtered under the same conditions as the variants in the presence of [ $\gamma$ - $^{32}$ P]ATP (Livesey & Martin, 1988). The phosphorylated proteins were stored on ice until ready to assay.

### Dephosphorylation of the luciferase

When the phosphorylated proteins were to be treated with alkaline phosphatase, the phosphorylation buffer contained no sodium glycerol 2-phosphate nor NaF. For the dephosphorylation reaction 0.7 unit of alkaline phosphatase/ $\mu$ l and 0.01 mM-protein kinase inhibitor (Cheng *et al.*, 1985) were added to the phosphorylation mixture.

### Effect of pH on activity and colour of the light emitted by the variants

Chemiluminescence from the enzymes was measured by diluting them 40-fold into an assay mix with pH ranging from 6 to 9 containing mixtures of 50 mM-Mes and 50 mM-Tris to give the desired pH, 0.3 mM-dithiothreitol, 0.2 mM-EDTA, 1 mg of BSA/ml, 12 mM-magnesium acetate, 0.2 mM-luciferin and 1.5 mM-ATP. Colour was assessed using a dual-wavelength luminometer fitted with narrow-band pass-interference filters, with a maximal transmission at 603 nm (red) and 545 nm (green) of 30.2 and 35.3 % respectively (Campbell *et al.*, 1985). The light produced by the luciferase reactions was measured simultaneously at the two wavelengths and the ratio of activity at 603 nm to activity at 545 nm was calculated. The ratio was corrected for the transmission of the filters, but not for the spectral sensitivity of the photomultiplier tubes, which at 603 nm was approximately 10 % of its value at 545 nm.

## RESULTS

### Characterization of PCR products

The PCR was used to amplify cDNA coding for wild-type firefly luciferase and for variants containing putative protein kinase A-recognition sites at position 217–220 (referred to as RRFS), kemptide at N-terminus (referred to as KNt) or kemptide at C-terminus (referred to as KCt).

The PCR products were characterized using three criteria: size on agarose-gel electrophoresis, formation of <sup>32</sup>P-labelled mRNA of the correct size on glyoxal/agarose-gel electrophoresis and translation *in vitro* of the mRNA to form light-emitting protein. This protein was compared with wild-type synthetic luciferase for molecular mass, specific activity, pH profile and colour and with firefly tails luciferase when appropriate. The PCR generated a single DNA band apparently of the correct predicted size for all the recombinant DNA, i.e. for wild-type and RRFS the predicted size is 1682 bp, for KCt the predicted size is 1703 bp, and for KNt a major band is present at the predicted size 1703 bp with a minor band at 380 bp (Fig. 1). The yields were 1–3 µg of DNA/0.1 ml of reaction mixture. No bands were seen without addition of primers or when template DNA was omitted.

Transcription of the PCR products with T7 RNA polymerase generated a major band of <sup>32</sup>P-labelled capped mRNA of the correct length, i.e. 1650 bp. Small quantities of longer and shorter mRNA products were observed (Fig. 2). The latter could not generate light-emitting protein because removal of 12 amino acids at the C-terminus reduces the activity by 99 % (Sala-Newby *et al.*, 1990b). The yields of capped mRNA were 4–8 molecules of RNA per DNA molecule, the lower yields corresponding to DNA coding for KNt. No mRNA was detected in gels when the DNA transcribed lacked T7 promoter in spite of the detection of [<sup>32</sup>P]UTP incorporation equivalent to 0.05 molecule of RNA per DNA molecule. mRNA generated light-emitting protein (Table 1) and a major <sup>35</sup>S-labelled protein band of the expected molecular mass (60 kDa) on SDS/PAGE (Fig. 3).

### Characterization of the recombinant variants

The new proteins were characterized using three criteria: specific activity (i.e. CL counts/µg of RNA and CL counts/ng of protein), effect of pH 6–9 on their activity and colour of the light emitted as assessed by the ratio of chemiluminescence at 603 nm to 543 nm.

The CL counts/10 s per ng of protein indicated the effect the modifications had on the catalytic activity of the protein. Luciferase with kemptide at the N- or C-terminus had a specific activity similar to that of the wild-type and extracted luciferase.

However, the specific activity of the RRFS variant was only 10–15 % of that of wild-type luciferase (Table 1). The specific chemiluminescent activity estimated per µg of RNA differed between the variants by a greater factor than the activity per µg

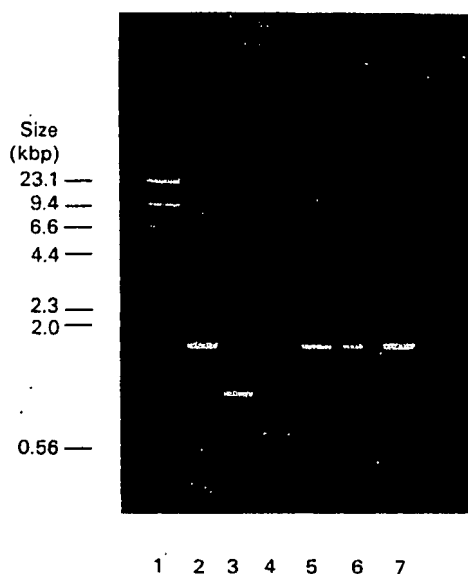


Fig. 1. Agarose-gel electrophoresis of cDNAs prepared by PCR

Wild-type firefly cDNA was amplified with oligonucleotides 105–100 (lane 2). The fragments of DNA that were used to prepare RRFS variant are shown in lanes 3 and 4. They correspond to PCR products obtained using oligonucleotide primers 107–100 (3' end) and 101–108 (5' end). The firefly RRFS cDNA is shown in lane 5; it was prepared by the amplification of DNA from lanes 3 and 4 in the presence of primers 105–100. cDNAs coding for variants with kemptide at N- and C-terminus are shown in lanes 6 and 7 respectively. Size markers were *Hind*III-digested λ DNA (lane 1).

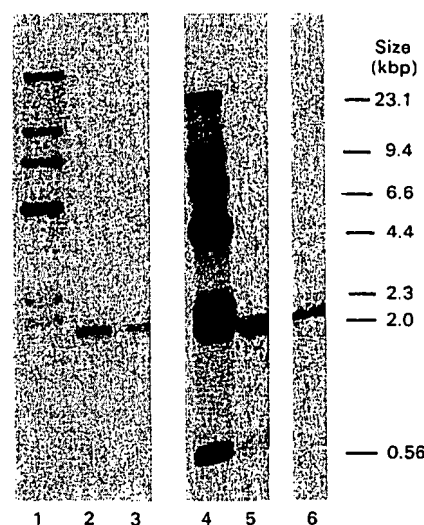


Fig. 2. Transcription products of the cDNAs

cDNAs produced by PCR were transcribed using T7 RNA polymerase and the <sup>32</sup>P-labelled mRNAs were separated by glyoxal/agarose-gel electrophoresis, dried and autoradiographed as described in the Materials and methods section. The size markers were <sup>32</sup>P-labelled *Hind*III-digested λ DNA (lanes 1 and 4) (Sambrook *et al.*, 1989). RNAs for the recombinant proteins are shown as follows: RRFS (lane 2), wild-type (lane 3), KNt (lane 5), KCt (lane 6).

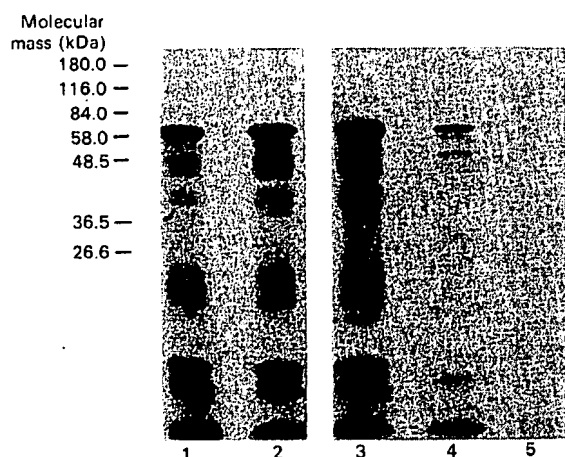


Fig. 3. Synthesis *in vitro* of recombinant proteins

mRNAs were translated using rabbit reticulocyte lysate in the presence of [ $^{35}$ S]methionine. The proteins were separated by SDS/PAGE under reducing conditions. RRFS (lane 1), wild-type (lane 2), KCt (lane 3), KNT (lane 4) and the products in the absence of mRNA (lane 5) are shown. Prestained molecular-mass (Da) markers were:  $\alpha_2$ -macroglobulin (180 000),  $\beta$ -galactosidase (116 000), fructose 6-phosphate kinase (84 000), pyruvate kinase (58 000), fumarase (48 500), lactate dehydrogenase (36 500) and triose phosphate isomerase (26 600).

Table 1. Specific activity of the luciferases

The values in parentheses indicate the number of independent DNA amplifications. Results are expressed as means  $\pm$  s.e.m. and when only two determinations were made the range is given.

Variant	CL counts/10 s per ng of protein	CL counts/10 s per $\mu$ g of RNA	Protein/RNA (mol/mol)
Wild-type	$(2.6 \pm 0.5) \times 10^5$ (3)	$(3.0 \pm 0.6) \times 10^7$ (5)	1.04
synthetic			
RRFS	$(3.0 \pm 1.0) \times 10^4$ (6)	$(3.9 \pm 1.3) \times 10^5$ (6)	0.11
KNT	$(2.1-3.7) \times 10^5$	$(4.2-4.7) \times 10^6$	0.14
KCt	$(2.1-2.0) \times 10^6$	$(3.7-8.9) \times 10^6$	0.27
Extracted	$2.1 \times 10^5$	—	—

of protein. All the variants with phosphorylation sites showed less activity per  $\mu$ g of RNA than the wild-type variant (Table 1). The number of molecules of protein produced per molecule of RNA in the translation assay, estimated from the two specific activities, confirmed that the normal synthetic wild-type enzyme yielded up to nine more copies of RNA than the other three (Table 1). The lower levels of translation shown could reflect differences in the secondary structure of the mRNA with the translation assay being optimized for the wild-type synthetic RNA. An additional effect due to a change in the codon usage cannot be ruled out. The RNA coding for KNT also contained an RNA band at approx. 500 bp (Fig. 2, lane 6) which would not translate into active protein as this would reduce the specific activity estimated per total RNA.

#### Effect of phosphorylation and dephosphorylation

Initial experiments using kemptide as a substrate for protein kinase A indicated that rabbit reticulocyte lysate inhibited phosphorylation of kemptide (results not shown). Gel filtration

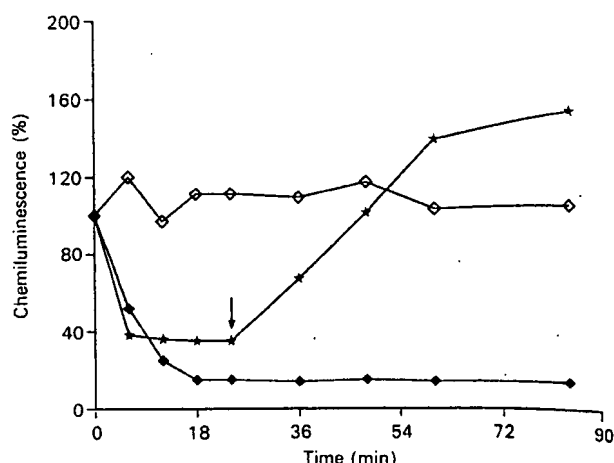


Fig. 4. Effect of phosphorylation-dephosphorylation on the activity of RRFS luciferase (V-217-R)

Partially purified variant RRFS was incubated at 30 °C as described in the Materials and methods section in the presence of kinase diluent only ( $\diamond$ ), protein kinase A catalytic subunit with ( $\blacklozenge$ ) and without ( $\star$ ) phosphate inhibitors. At 24 min alkaline phosphatase and protein kinase A inhibitor were added ( $\downarrow$ ). Samples were taken from the tubes at various times up to 90 min, diluted immediately 40-fold into luciferase assay mixture pH 7.2 and the chemiluminescence was measured for 10 s. Activity at time 0 was measured before addition of kinase. Results are presented as percentage of activity at time 0. Each point is a mean of two and a representative experiment is shown. Experiments were carried out with protein produced from two separate PCRs.

removed the inhibitory activity and  $1.7 \pm 0.2$  ( $n = 3$ ) nmol of phosphate ( $1.3-1.4$  mol when phosphatase inhibitors were omitted) was incorporated into kemptide after 20 min incubation per 40  $\mu$ l of reaction mixture.

Incubation of the RRFS luciferase variant with protein kinase A catalytic subunit in the presence of ATP resulted in a decrease in its catalytic activity to  $19 \pm 4\%$  ( $n = 5$ ) within 20 min, and remained at this level for the duration of the experiment, i.e. 90 min (Fig. 4). When alkaline phosphatase was added to the phosphorylated RRFS luciferase, the chemiluminescent activity increased to control levels within 30 min. No effect of protein kinase A was observed on the activity of wild-type luciferase, nor on recombinant luciferases with kemptide at the *N*- or *C*-terminus, at any pH (Figs. 5a and 5b).

Attempts to demonstrate a change in pI between the various recombinant luciferases, using isoelectric focusing, were unsuccessful, because of artifactual bands generated from the focusing procedure. However, the major band for recombinant and extracted luciferase had the same pI (6.6).

Recombinant wild-type luciferase had a pH optimum of around 7.8, identical with that of the extracted luciferase (Fig. 6a). Addition of kemptide at the *N*- or *C*-terminus appeared to have no effect on the pH profile (Figs. 5b and 6a). Similarly these three recombinant proteins had similar colour shifts to the red at acidic pH (Fig. 6b). In contrast the RRFS mutant luciferase showed both an altered pH profile with optimum activity at pH 7.2 (Figs. 5a and 6a) and a shift in colour to the green at acid pH (Fig. 6b). The inhibitory effect of phosphorylation on RRFS activity was most marked at its optimum pH (Fig. 5a). The ratio of light emission at 603 nm/543 nm measured at pH 7.5 changed from 0.16 to 0.32 after phosphorylation, indicating that the light became redder. Since the detection system used for activity measurements was less sensitive to red light, this red shift may

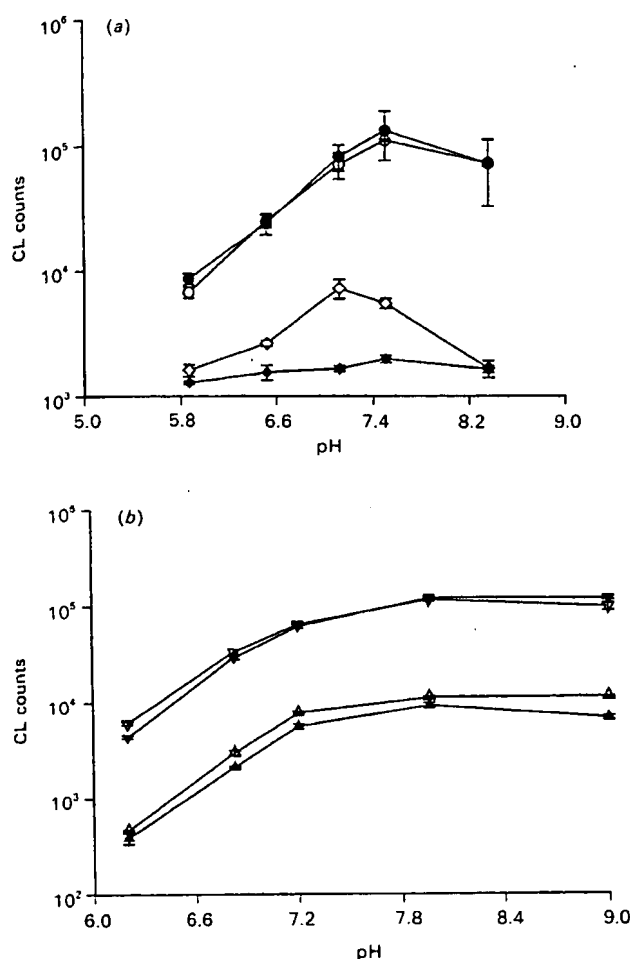


Fig. 5. Effect of pH and protein kinase A on the activity of recombinant luciferases

The variants were incubated for 15 min in the presence of protein kinase A (closed symbols) or kinase diluent only (open symbols) as described in the Materials and methods section. The resulting enzyme activity was then measured in triplicate (mean  $\pm$  S.E.M.) at various pH values. (a) Wild-type (●, ○) and RRFS (◆, ◇). (b) KNt (▲, △) and KCt (▼, ▽).

partly explain the decrease in activity for the phosphorylated enzyme.

## DISCUSSION

The results presented demonstrate that DNA amplification coupled to transcription-translation *in vitro* allowed the generation and characterization of firefly luciferase variants containing phosphorylation sites. Only one of the variants (RRFS) showed a decrease in its activity when incubated with the catalytic subunit of protein kinase A in the presence of ATP, and the effect was reversed by addition of alkaline phosphatase (Figs. 4, 5a and 5b). The enzyme activity per unit of protein of the wild-type variant and the luciferases with kemptide at the N- or C-terminus were indistinguishable from that of the extracted luciferase (Table 1). The activity expressed per unit of RNA was more variable and lower for all the variants with phosphorylation sites than for the wild-type. The pH-activity profile for KNt, KCt and wild-type were very similar, but RRFS had a lower pH optimum (Fig. 6a). The colour of the light emitted was assessed by

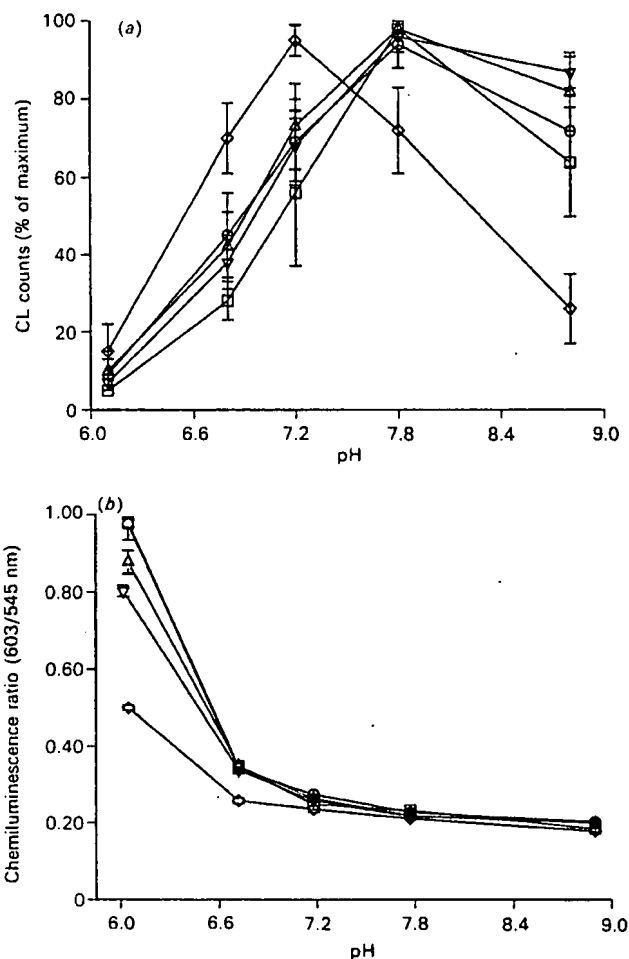


Fig. 6. Effect of pH on the activity and colour of the light emitted by luciferase variants

(a) pH optimum curve. The results are expressed as % of maximal activity (mean  $\pm$  S.E.M.) from three to five experiments, each in triplicate. ○, Wild-type recombinant; □, extracted luciferase; ◇, RRFS; △, KNt; ▽, KCt. (b) pH effect on the colour of the light produced. The ratio of chemiluminescent counts at 603 nm and 545 nm was measured in triplicate at each pH (mean  $\pm$  S.E.M.). ○, Wild-type recombinant; □, extracted luciferase; ◇, RRFS; △, KNt; ▽, KCt. The luciferase from firefly tails was purified as described by Sala-Newby *et al.* (1990b).

measuring the ratio of activities at 603 nm and 545 nm (Fig. 6b). At alkaline pH no significant differences were detected but as the pH was decreased the variant RRFS had a significantly lower ratio, indicating that the light emitted was greener than for all the others. As chemiluminometers contain photomultipliers which are more sensitive to green than red light this colour change cannot explain the decrease in specific activity measured. The activity was measured under saturating concentrations of ATP and luciferin, suggesting that the  $V_{max}$  was decreased.

Several beetle luciferases have now been cloned: *Photinus pyralis*, *Pyrophorus plagiophthalmus* and *Luciola cruciata* (de Wet *et al.*, 1987; Wood *et al.*, 1989a,b; Tatsumi *et al.*, 1989). Spectral changes are known to occur in the light emitted by firefly luciferase in response to changes in pH and temperature, and in the presence of heavy metals (Seliger & McElroy, 1964). Work on four click-beetle luciferases that show 94-99% sequence homology demonstrated that a small number of amino acid



substitutions were responsible for the different colours displayed by the luciferases. The spectral shift between luciferases yellow-green and yellow belong to the amino acid set, R-223, L-238 → E, V with the effect probably being due to R-223 → E (Wood *et al.*, 1989a,b). Since all the beetle luciferases use the same luciferin, the colour of the light emitted in the reaction must depend on the environment around the emitter (i.e. oxyluciferin). The oxyluciferin can exist as a monoanion (ketonic form) or dianion (enolic form) at acid and basic pH respectively. The presence of an arginine in position 223 of the click-beetle yellow-green luciferase seemed to be responsible for a shift to the green in the light it emitted. The change V-217 → R-217 that generated RRFS in *Photinus* luciferase introduced a basic amino acid in that area of the protein and also resulted in a shift to the green of the light emitted, suggesting that a positive charge there stabilized the oxyluciferin dianionic form, the green emitter.

The phosphorylation of the RRFS variant by the catalytic subunit of protein kinase A decreased its activity, and dephosphorylation reversed the effect. The decrease in activity was accompanied by a spectral shift to the red that can account for part of the lower activity measured. The other two variants, KNt and KCt, showed no detectable differences from the wild-type luciferase in any aspect. The luciferase with kemptide at the C-terminus was expected to show properties different from the normal luciferase in view of the fact that the removal of 12 amino acids at the C-terminus nearly abolishes activity (Sala-Newby *et al.*, 1990b). The removal of three amino acids (results not shown) and the addition of the seven amino acids from kemptide at the C-terminus did not affect the catalytic properties. This could be important when using firefly luciferase or its variant in eukaryotic cells because the last three amino acids of the C-terminus contain a peroxisomal targeting signal (Keller *et al.*, 1987; Gould *et al.*, 1987).

The RRFS variant provides, for the first time, an indicator potentially useful for measuring protein phosphorylation in intact cells, and has also highlighted a domain within the enzyme that results in changes in colour in response to a change in charge. Recognition peptides for other kinases could thus be engineered in this region of the protein, thereby establishing a universal strategy for measuring any protein kinase and visualizing it in living cells (Hooper *et al.*, 1990).

We thank Mr. Andrew Trimby for technical assistance, Dr. Rose Ann Padua for the synthesis of the oligonucleotides and both her and Dr. K. Murray for helpful discussions. This work was funded by the Medical Research Council and the Agriculture and Food Research Council.

## REFERENCES

- Berridge, M. J. & Irvine, R. F. (1989) *Nature* (London) **341**, 197–205  
 Campbell, A. K. (1983) *Intracellular Calcium: Its Universal Role as Regulator*, Wiley, Chichester

- Campbell, A. K. (1988) *Chemiluminescence: Principles and Applications in Biology and Medicine*, Horwood/VCH, Chichester and Weinheim  
 Campbell, A. K. (1989) Br. Pat. Appl. 8916806  
 Campbell, A. K. (1990) in *Oxygen Radicals and Cellular Damage* (Duncan, C. J., ed.), pp. 189–217, Cambridge University Press, Cambridge  
 Campbell, A. K., Roberts, P. A. & Patel, A. (1985) in *Alternative Immunoassays* (Collins, W. P., ed.), pp. 153–183, Wiley, Chichester  
 Cheng, H. C., Van Patten, S. M., Smith, A. J. & Walsh, D. A. (1985) *Biochem. J.* **231**, 655–661  
 Cobbold, P. H. & Rink, T. J. (1987) *Biochem. J.* **248**, 313–328  
 Cohen, P. (1988) *Proc. R. Soc. London Ser. B* **234**, 115–144  
 Davies, E. V., Hallett, M. B., Williams, B. D. & Campbell, A. K. (1991) *Immunology* **73**, 228–234  
 de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. (1987) *Mol. Cell. Biol.* **7**, 725–737  
 Gould, S. J., Keller, G. A. & Subramani, S. (1987) *J. Cell Biol.* **105**, 2923–2931  
 Hallett, M. B., Davies, E. V. & Campbell, A. K. (1990) *Cell Calcium* **11**, 655–663  
 Higuchi, R. (1990) in *PCR protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J., eds.), pp. 177–183, Academic Press, San Diego and London  
 Hooper, C. E., Ansorge, R. E., Browne, H. M. & Tomkins, P. (1990) *J. Biol. Chem.* **265**, 123–130  
 Jenkins, T., Sala-Newby, G. & Campbell, A. K. (1990) *Biochem. Soc. Trans.* **18**, 463–464  
 Keller, G. A., Gould, S., DeLuca, M. & Subramani, S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3264–3268  
 Kemp, B. E., Graves, D. J., Benjamini, E. & Krebs, E. G. (1977) *J. Biol. Chem.* **252**, 4888–4894  
 Laemmli, U. K. (1970) *Nature* (London) **227**, 680–685  
 Li, B. L., Langer, J. A., Schwartz, B. & Pestka, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 558–562  
 Livesey, S. A. & Martin, T. J. (1988) *Methods Enzymol.* **159**, 105–118  
 Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275  
 Murray, K., England, P. J., Lynham, J. A., Mills, D., Schmitz-Peiffer, C. & Reeves, M. L. (1990) *Biochem. J.* **267**, 703–708  
 Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491  
 Sala-Newby, G., Kalsheker, N. & Campbell, A. K. (1990a) *Biochem. Soc. Trans.* **18**, 459–460  
 Sala-Newby, G., Kalsheker, N. & Campbell, A. K. (1990b) *Biochem. Biophys. Res. Commun.* **172**, 477–482  
 Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY  
 Seliger, H. H. & McElroy, W. D. (1964) *Proc. Natl. Acad. Sci. U.S.A.* **52**, 81–85  
 Stoffel, E. S., Koeberl, D. D., Sarkar, G. & Sommer, S. S. (1988) *Science* **239**, 491–494  
 Tatsumi, H., Masuda, T., Kajiyama, N. & Nakano, E. (1989) *J. Biol. Chem.* **264**, 75–78  
 Wood, K. V., Lam, Y. A., McElroy, W. D. & Seliger, H. H. (1989a) *J. Biol. Chem.* **264**, 31–39  
 Wood, K. V., Lam, Y. A. & McElroy, W. D. (1989b) *Science* **244**, 700–702  
 Zetterqvist, O., Ragnarsson, U., Humble, E., Berglund, L. & Engstrom, L. (1976) *Biochem. Biophys. Res. Commun.* **70**, 696–703

Received 25 April 1991/11 June 1991; accepted 18 June 1991

---

**The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications**

---

Paul M.Sharp<sup>1,2\*</sup> and Wen-Hsiung Li<sup>2</sup>

---

<sup>1</sup>Department of Genetics, Trinity College, Dublin 2, Ireland and <sup>2</sup>Center for Demographic and Population Genetics, University of Texas, PO Box 20334, Houston, TX 77225, USA

---

Received August 4, 1986; Revised and Accepted December 18, 1986

---

**ABSTRACT**

A simple, effective measure of synonymous codon usage bias, the Codon Adaptation Index, is detailed. The index uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a score for a gene is calculated from the frequency of use of all codons in that gene. The index assesses the extent to which selection has been effective in moulding the pattern of codon usage. In that respect it is useful for predicting the level of expression of a gene, for assessing the adaptation of viral genes to their hosts, and for making comparisons of codon usage in different organisms. The index may also give an approximate indication of the likely success of heterologous gene expression.

**INTRODUCTION**

The determination of the DNA sequences of a large number of genes from a wide variety of species has revealed that, in a large proportion of cases, the alternative synonymous codons for any one amino acid are not used randomly (1, and references therein). Further, it has been noted that a part of this nonrandom usage is species, or rather taxon, specific (2). However, within species there is considerable heterogeneity between genes, and in the two best studied organisms, namely *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, there is a clear positive correlation between degree of codon bias and level of gene expression (3,4). Examination of large data sets from these species reveals that within species differences are largely in the degree rather than the direction of codon usage bias (5,6).

For many reasons it is desirable to quantify the degree of bias in codon usage in each gene in such a way that comparisons can be made both within and between species. One approach to this problem is to devise a measure for assessing the degree of deviation from a postulated impartial pattern of usage. The codon preference bias proposed by McLachlan et al. (7) is such a measure. Recently Sharp et al. (5) have proposed to calculate the chi square value for the deviation from random codon usage and then scale

the value by the gene length (number of codons) so that comparisons can be made between genes.

Another approach is to assess the relative merits of different codons from the viewpoint of translational efficiency. For example, Ikemura (1,8,9) has identified certain "optimal" codons in E.coli and yeast which are expected to be translated more efficiently than others, and calculated the frequency of optimal codons in a gene. The codon bias index of Bennetzen and Hall (4), for use with yeast genes, is essentially similar. Such indices are certainly useful, but have several disadvantages. First, some amino acids are usually excluded because it is not clear which codons are "optimal". Second, all codons considered are classified into only two categories, i.e., optimal and nonoptimal, with no recognition that some codons within each category are better than others. Third, there is no good basis for comparison between species because the proportional division of the codon table into the two categories may differ; e.g., Ikemura (1) identified 21 optimal codons for 14 amino acids in E.coli, and 19 optimal codons for 13 amino acids in yeast.

Gribskov et al. (10) have recently proposed another index, the codon preference statistic. This statistic is based on the ratio of the likelihood of finding a particular codon in a highly expressed gene to the likelihood of finding that codon in a random sequence with the same base composition as that in the sequence under study. They show that the statistic is useful for locating genes in sequenced DNA, for predicting the relative level of their expression, and for detecting sequencing errors. However, the statistic is not normalized and therefore the values for two genes encoding proteins with different amino acid compositions can be quite different even if both genes use only the "best" codons.

With various purposes in mind we have devised a new index. It is similar to the codon preference statistic but is normalized so that it is convenient for making comparisons both within and between species. After describing the index, we show some rather varied applications and indicate certain advantages over other indices. In recognition of the role of natural selection in producing high levels of codon bias, we call this statistic the Codon Adaptation Index.

### METHODS

We recognize that even in E.coli and yeast the factors determining the frequency of synonymous codon usage are not completely understood, but that

several points are clear: the pattern of codon usage in any particular gene is largely determined by natural selection and mutation (5,6); selection appears to occur via translational efficiency, so that synonymous codon usage in highly expressed genes is under the strongest selective constraints (4,8,9); in *E.coli* and yeast, very highly expressed genes appear to have the greatest degree of synonymous codon bias (3-6,8). From these points it is deduced that the pattern of codon usage in very highly expressed genes can reveal (i) which of the alternative synonymous codons for an amino acid is the most efficient for translation, and (ii) the relative extent to which other codons are disadvantageous.

The first step is, then, to construct a reference table of relative synonymous codon usage (RSCU) values from very highly expressed genes of the organism in question. An RSCU value for a codon is simply the observed frequency of that codon divided by the frequency expected under the assumption of equal usage of the synonymous codons for an amino acid (5). Thus,

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}} \quad [1]$$

where  $X_{ij}$  is the number of occurrences of the  $j$ th codon for the  $i$ th amino acid, and  $n_i$  is the number (from one to six) of alternative codons for the  $i$ th amino acid. The relative adaptiveness of a codon,  $w_{ij}$ , is then the frequency of use of that codon compared to the frequency of the optimal codon for that amino acid:

$$w_{ij} = RSCU_{ij} / RSCU_{imax} = X_{ij} / X_{imax} \quad [2]$$

where  $RSCU_{imax}$  and  $X_{imax}$  are the RSCU and  $X$  values for the most frequently used codon for the  $i$ th amino acid.

Codon usage data have been compiled previously for 165 genes from *E.coli* (6), and for 110 genes from yeast (5). To obtain reference RSCU values, we have taken the 27 very highly expressed *E.coli* genes compiled by Sharp and Li (6), which include genes encoding 17 ribosomal proteins, four outer membrane proteins and four elongation factors. For yeast a set of 24 genes has been taken from the high expression group previously identified (5). These include 16 genes encoding ribosomal proteins, one for an elongation factor, and seven loci encoding very abundant enzymes. The RSCU

Table 1. Values of RSCU and w for codons in very highly expressed genes from *E.coli* and yeast.

		<u>E.coli</u>		Yeast				<u>E.coli</u>		Yeast	
		RSCU	w	RSCU	w			RSCU	w	RSCU	w
Phe	UUU	0.456	0.296	0.203	0.113	Ser	UCU	2.571	1.000	3.359	1.000
	UUC	1.544	1.000	1.797	1.000		UCC	1.912	0.744	2.327	0.693
Leu	UUA	0.106	0.020	0.601	0.117		UCA	0.198	0.077	0.122	0.036
	UUG	0.106	0.020	5.141	1.000		UCG	0.044	0.017	0.017	0.005
Leu	CUU	0.225	0.042	0.029	0.006	Pro	CCU	0.231	0.070	0.179	0.047
	CUC	0.198	0.037	0.014	0.003		CCC	0.038	0.012	0.036	0.009
	CUA	0.040	0.007	0.200	0.039		CCA	0.442	0.135	3.776	1.000
	CUG	5.326	1.000	0.014	0.003		CCG	3.288	1.000	0.009	0.002
Ile	AUU	0.466	0.185	1.352	0.823	Thr	ACU	1.804	0.965	1.899	0.921
	AUC	2.525	1.000	1.643	1.000		ACC	1.870	1.000	2.063	1.000
	AUA	0.008	0.003	0.005	0.003		ACA	0.141	0.076	0.025	0.012
Met	AUG	1.000	1.000	1.000	1.000		ACG	0.185	0.099	0.013	0.006
Val	GUU	2.244	1.000	2.161	1.000	Ala	GCU	1.877	1.000	3.005	1.000
	GUC	0.148	0.066	1.796	0.831		GCC	0.228	0.122	0.948	0.316
	GUA	1.111	0.495	0.004	0.002		GCA	1.099	0.586	0.044	0.015
	GUG	0.496	0.221	0.039	0.018		GCG	0.796	0.424	0.004	0.001
Tyr	UAU	0.386	0.239	0.132	0.071	Cys	UGU	0.667	0.500	1.857	1.000
	UAC	1.614	1.000	1.868	1.000		UGC	1.333	1.000	0.143	0.077
ter	UAA	--	--	--	--	ter	UGA	--	--	--	--
ter	UAG	--	--	--	--	Trp	UGG	1.000	1.000	1.000	1.000
His	CAU	0.451	0.291	0.394	0.245	Arg	CGU	4.380	1.000	0.718	0.137
	CAC	1.549	1.000	1.606	1.000		CGC	1.561	0.356	0.008	0.002
Gln	CAA	0.220	0.124	1.987	1.000		CGA	0.017	0.004	0.008	0.002
	CAG	1.780	1.000	0.013	0.007		CGG	0.017	0.004	0.008	0.002
Asn	AAU	0.097	0.051	0.100	0.053	Ser	AGU	0.220	0.085	0.070	0.021
	AAC	1.903	1.000	1.900	1.000		AGC	1.055	0.410	0.105	0.031
Lys	AAA	1.596	1.000	0.237	0.135	Arg	AGA	0.017	0.004	5.241	1.000
	AAG	0.404	0.253	1.763	1.000		AGG	0.008	0.002	0.017	0.003
Asp	GAU	0.605	0.434	0.713	0.554	Gly	GGU	2.283	1.000	3.898	1.000
	GAC	1.395	1.000	1.287	1.000		GGC	1.652	0.724	0.077	0.020
Glu	GAA	1.589	1.000	1.968	1.000		GGA	0.022	0.010	0.009	0.002
	GAG	0.411	0.259	0.032	0.016		GGG	0.043	0.019	0.017	0.004

## Genes used:

*E.coli* - 17 ribosomal protein genes, 4 elongation factor genes, 4 outer membrane protein genes, *recA*, *dnaK* (data from Ref.6)

Yeast - 16 ribosomal protein genes, *TEF 1*, 2 enolase genes, 2 GA-3-PDH genes, *ADH 1*, *PGK*, pyruvate kinase (data sources given in Ref.5)

and  $w$  values obtained for very highly expressed genes from E.coli and yeast are given in Table 1.

The Codon Adaptation Index (CAI) for a gene is then calculated as the geometric mean of the RSCU values (from Table 1) corresponding to each of the codons used in that gene, divided by the maximum possible CAI for a gene of the same amino acid composition, i.e.,

$$CAI = CAI_{obs} / CAI_{max} \quad [3]$$

where

$$CAI_{obs} = \left( \prod_{k=1}^L RSCU_k \right)^{1/L} \quad [4]$$

$$CAI_{max} = \left( \prod_{k=1}^L RSCU_{kmax} \right)^{1/L} \quad [5]$$

where  $RSCU_k$  is the RSCU value for the  $k$ th codon in the gene,  $RSCU_{kmax}$  is the maximum RSCU value for the amino acid encoded by the  $k$ th codon in the gene, and  $L$  is the number of codons in the gene.

Note that if a certain codon is never used in the reference set then the CAI for any other gene in which that codon appears becomes zero. To overcome this problem we assign a value of 0.5 to any  $X_{ij}$  that would otherwise be zero. Also, the number of AUG and UGG codons are subtracted from  $L$ , since the RSCU values for AUG and UGG are both fixed at 1.0, and so do not contribute to the CAI.

As illustration, consider the rpsU gene from E.coli which, excluding the initiation codon, comprises 70 codons and has the sequence:

.CCG.GTA.ATT.AAA.GTA. . . . .

For that sequence and from the RSCU values in Table 1:

$$CAI_{obs} = (3.288 \times 1.111 \times 0.466 \times 1.596 \times 1.111 \times \dots)^{1/70}$$

$$\text{and } CAI_{max} = (3.288 \times 2.244 \times 2.525 \times 1.596 \times 2.244 \times \dots)^{1/70}$$

From these two values and equation [3] we can obtain the CAI value.

We note that equation [3] is exactly equivalent to:

$$CAI = \left( \prod_{k=1}^L w_k \right)^{1/L} \quad [6]$$

Table 2. CAI values for E.coli and yeast genes.

<u>E.coli</u>		yeast	
gene	CAI	gene	CAI
17 RPs	0.467-0.813	16 RPs	0.529-0.915
<u>rpsU</u>	0.726	histones	0.532-0.733
<u>rpoD</u>	0.582		
<u>dnaG</u>	0.271	2u plasmid	0.099-0.106
<u>lacI</u>	0.296	<u>GAL 4</u>	0.116
<u>trpR</u>	0.267	<u>PPR 1</u>	0.114
<u>lpp</u>	0.849 <sup>a</sup>	<u>GPD 1</u>	0.929 <sup>a</sup>
<u>hsdS</u>	0.218 <sup>b</sup>	<u>mat A2</u>	0.098 <sup>b</sup>

RPs - ribosomal protein genes.

a highest CAI value among data set.

b lowest CAI value among data set.

where  $w_k$  is the  $w$  value for the  $k$ th codon in the gene (see equation [2]).

Therefore, for rpsU:

$$\text{CAI} = (1.00 \times 0.495 \times 0.185 \times 1.000 \times 0.495 \times \dots)^{1/70}$$

Equation [6] saves computation time. To overcome real number underflow problems in computer calculations, equation [6] can be computed as:

$$\text{CAI} = \exp \frac{1}{L} \sum_{k=1}^L \ln w_k \quad [7]$$

or from a codon usage table:

$$\text{CAI} = \exp \frac{1}{L} \sum_{i=1}^{18} \sum_{j=1}^{n_i} X_{ij} \ln w_{ij} \quad [8]$$

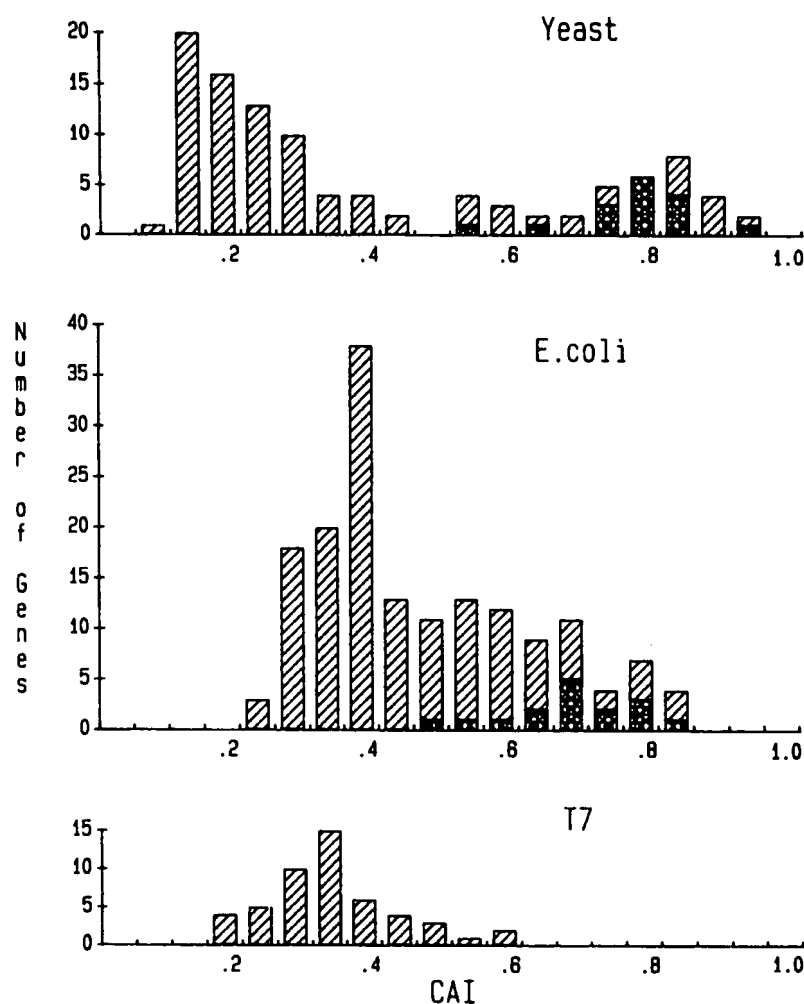
where  $X_{ij}$  and  $n_i$  are as defined in equation [1].

There is no intrinsic effect of gene length ( $L$ ) on CAI, but CAI values from short genes may be more variable due to sampling effects.

#### APPLICATIONS and DISCUSSION

##### Predicting levels of gene expression within a species.

CAI values clearly parallel levels of gene expression. Ribosomal protein genes are highly expressed, and have generally high CAI values



**Figure 1.** Distribution of CAI values for (a) 106 yeast genes, (b) 165 *E. coli* genes, and (c) 50 bacteriophage T7 genes. In (a) and (b) ribosomal protein genes are cross-hatched. Plasmid genes are excluded.

(Table 2, Figure 1). Among yeast ribosomal protein genes only that encoding S33 has a CAI < 0.6, and it is a very short gene (L = 65). Lowly expressed regulatory genes (e.g., *lacI*, *trpR* in *E. coli*; *GAL 4*, *PPR 1* in yeast) have low CAI values (Table 2). In *E. coli* the relationship between codon bias and gene expression is perhaps best illustrated by considering operons (as suggested by Gouy and Gautier, Ref.3). For example, within the macromolecular synthesis operon the expression levels are *rpsU* >> *rpoD* >> *dnaG* (11), and the CAI values for these genes are 0.726, 0.582 and 0.271, respectively (Table 2). Eight of the nine genes of the *unc* operon encode the



Table 3. CAI values for genes in the unc operon of E.coli.

Pos	Gene	CAI	L	Gene name	Product amount	sector
1	<u>papI</u>	0.238	127		??	
2	<u>papD</u>	0.400	253	chi	1	F <sub>0</sub>
3	<u>papH</u>	0.583	71	omega	10	
4	<u>papF</u>	0.482	152	psi	2	
5	<u>papE</u>	0.374	169	delta	1	F <sub>1</sub>
6	<u>papA</u>	0.665	501	alpha	3	
7	<u>papC</u>	0.403	273	gamma	1	
8	<u>papB</u>	0.650	444	beta	3	
9	<u>papG</u>	0.474	133	epsilon	1	

Pos : gene position within the operon (1 = 5').  
 The relative amount of each gene product in the ATPase complex is taken from Ref.12.

eight subunits of the F<sub>0</sub> and F<sub>1</sub> sectors of the H<sup>+</sup>-ATPase complex, and the stoichiometry of these subunits is known (12). The CAI value is clearly correlated with the level of gene expression among the genes encoding subunits of the F<sub>1</sub> sector (Table 3), with the CAI values for papA and papB being similar, and much higher than those for papE, papC and papG. Among genes encoding subunits in the F<sub>0</sub> sector the rank order of CAI values corresponds to the relative amounts of the gene products required. The CAI for papH is perhaps surprisingly low, but this is a very short gene (Table 3). The function of papI is unknown. The CAI value for papI is very low, and may indicate that this is a regulatory gene, or perhaps (see below) a noncoding open reading frame.

Although many of the measures of codon bias discussed in the Introduction seem to be positively correlated with gene expression, we feel that CAI has the twin advantages of being simple to calculate and making greater quantitative use of available information (see 'Comparison of CAI with other indices' below).

The positive correlation between degree of synonymous codon bias and expression level in E.coli (and yeast) seems firmly established, but the causal relationship between the two has been debated. We have concluded elsewhere (6) that the degree of codon bias reflects the past action of natural selection -- it is indicative of the level at which the gene is expressed, rather than dictating that level. This seems to concur with conclusions drawn from a theoretical model of the translation process (13).

Table 4. CAI values for mammalian genes using E.coli and yeast RSCU values.

Heterologous gene	Host	
	<u>E.coli</u>	Yeast
Human alpha interferon	0.218	0.099
Human insulin	0.307	0.043
Human growth hormone	0.287	0.082
Human factor VIII	0.205	0.114
Human factor IX	0.263	0.176
Bovine chymosin	0.326	0.086

Predicting levels of heterologous gene expression.

There is experimental evidence that certain codons can affect expression level (14-17). For example, the AGG codon markedly affects the translation rate of genes in E.coli (14,15). This suggests that for a heterologous gene to have a maximal level of expression its codon usage must correspond to that of the host. By using the RSCU values of potential hosts to calculate CAI values for a heterologous gene it should be possible to predict how well suited that gene would be to the translational systems of those hosts. In Table 4 the CAI values of some genes of biotechnological interest are given for two different potential hosts, E.coli and yeast. In each case these mammalian genes seem better 'adapted' to E.coli, suggesting that high expression might be more easily obtained in that system. Of course, in reality, the choice of host would probably depend on other practicalities. The CAI would, however, suggest whether it is likely to be either necessary or of any benefit to chemically synthesize a new gene, to include more appropriate codons. It should be stressed that the CAI is only an approximate indication of the suitability of the codon usage within a gene. For example, it takes no account of the distribution of codons along the gene, yet theoretical considerations suggest that this may be very important (18).

A measure of evolutionary adaptedness.

Under certain natural circumstances foreign genes are expressed in host organisms. Viral genes are an obvious example. Codon usage in the many bacteriophages which do not encode their own tRNA molecules should be adapted to the translational machinery of the host. Then the CAI, using host RSCU values, is an estimate of the degree of adaptation. For example, comparison of the pattern of codon usage in the genes of bacteriophage T7

Table 5. CAI values for homologous genes from E.coli and T7.

<u>E.coli</u> gene	CAI	T7 gene	CAI
<u>ssb</u>	0.605	2.5	0.573
<u>dnaG</u>	0.271	4	0.301
<u>polA</u>	0.391	5	0.341
		6	0.387

with the relative abundance of cognate tRNA molecules in E.coli (considered to be the usual host of T7) suggests that T7 genes are not so well adapted as E.coli's own genes, although there is clearly some adaptation (19,20). This seems to be confirmed by contrasting the distribution of CAI values for T7 genes with those of E.coli (Figure 1). However, the difference seen in Figure 1 could arise in part because the genes contrasted encode different products; for example, T7 encodes no ribosomal proteins. It has been reported that four genes in T7 are homologous to three E.coli genes (21). A comparison of these genes (Table 5) is not conclusive, because only ssb is highly adapted in E.coli, although in that case the T7 gene does have a lower CAI. The four T7 genes as a group do not seem to be significantly less adapted than the three E.coli genes.

In cases where it has not been clear which organism represents the major host for a virus it may prove informative to calculate CAI values with the different RSCU values of potential hosts. For example, despite approximately 65% DNA homology between ØX174 and G4, the genomes of these two "coliphages" show a remarkable difference with respect to the frequency of the recognition sites of enterobacterial restriction enzymes (22). While ØX174 (as well as several other coliphages) has a significant avoidance of these sites, presumably reflecting adaptation to infecting E.coli, G4 does not. However, CAI values for the 10 genes of ØX174 and G4 are very similar, suggesting that the patterns of codon usage of the two phages are adapted (to E.coli) to equivalent extents.

Natural foreign gene expression would also occur if genes undergo horizontal transfer. Felmlee et al. (23) have discussed a possible example. They reported the DNA sequence of a region of the E.coli chromosome encoding four hemolysin genes, and found that their base composition and codon usage are atypical of that species. This, together with the observation that these genes are found in only a limited number of E.coli strains, was taken as

evidence that the genes represent a recent acquisition to this species (23). The CAI values for these genes are indeed very low, ranging from 0.202 to 0.243. These values are lower than those for nearly all other E.coli genes (see Figure 1, in which the hemolysin genes are not included), including some (e.g., araC and dnaG) which are expressed at very low levels. Hemolysin is an extracellular protein and would be expected to be expressed at much higher levels than araC or dnaG, so these low CAI values suggest that the hemolysin genes are not well adapted to E.coli, and seem to confirm the suggestion of a recent acquisition. If reference RSCU data were available for a variety of organisms from which the genes could have been transferred, it might be possible to determine the most likely source by comparison of CAI values.

If plasmids were regularly subject to interspecific transfer, then their genes might not become adapted to any one host. Genes on E.coli plasmids tend to have less codon bias than chromosomal genes (3). We note that the three genes of the yeast 2 micron plasmid have very low CAI values (Table 2).

#### Synonymous codon usage and the rate of molecular evolution.

A major prediction of the neutral theory of molecular evolution (24) is an inverse relationship between the rate of evolution and the degree of selective constraint, i.e., the stronger the constraint the slower the rate of molecular evolution. Indeed, a great deal of evidence confirms this, including the observation that pseudogenes, which are under no apparent constraint, are the fastest evolving DNA sequences (25). That synonymous substitutions in protein coding genes occur at a slower rate than substitutions in pseudogenes (26,27) implies that there are selective constraints on the former. If the differences between genes in degree of codon usage bias largely reflect differences in selection pressure on synonymous codons, then the rate of synonymous substitution would be inversely related to the degree of codon bias. The CAI can be used to quantify this relationship. Comparisons of E.coli and Salmonella typhimurium genes do indeed show a significant negative correlation between the rate of synonymous substitution and the CAI (28).

#### Comparison of codon usage in different organisms.

Meaningful comparisons of codon usage in different organisms can be made if care is taken in defining the reference set of genes from which the RSCU values are calculated. The reference sets we have chosen for E.coli and yeast comprise very similar collections of genes, yet the distribution of

CAI values for genes from these two organisms are rather different. Very highly expressed genes in yeast have on average a more extreme codon bias than their counterparts in E.coli, as seen for example with ribosomal protein genes (Table 2). The reference set of RSCU values reflects this, and so the genes with least codon usage bias in yeast have lower CAI values than genes in E.coli, as a result. It is particularly interesting to note that cluster analysis of yeast genes based on their synonymous codon usage clearly differentiates two groups, identified as comprising highly and moderately/lowly expressed genes (5), and that those two groups correspond almost exactly to the bimodal distribution of CAI values for yeast genes in Figure 1. By contrast, cluster analysis does not so easily differentiate highly and lowly expressed genes in E.coli or in T7 (5) and the distributions of CAI values from those organisms are unimodal (Figure 1). It is not clear why selection has apparently been more successful in producing high codon bias in yeast than in E.coli. Li (29) has shown that the effectiveness of selection in maintaining synonymous codon bias depends largely on the strength of selection and effective population size. It could be that the strength of selection is stronger in yeast than in E.coli because the required amount of certain gene products, such as ribosomal proteins, is larger. It is also possible that the effective population size is larger in yeast than in E.coli because the latter has a largely clonal population structure (30).

We note that comparisons between species can be difficult when the reference sets of genes have quite different levels of bias in codon usage. For example, very highly expressed genes have a much lower bias in codon usage in Bacillus subtilis than in E.coli or yeast (Shields and Sharp, in prep.). Then, in B.subtilis, there are few codons with very low  $w$  values. As a consequence, CAI values for other genes in B.subtilis are, on average, higher than those seen in the other species, even though the B.subtilis genes have clearly less bias. The  $CAI_{obs}$  given by equation [4] is less affected by this difference in the reference set, and may form a better basis for comparison between species under these circumstances.

## Identification of protein-coding reading frames.

Several of the indices of codon usage bias were originally devised in order to ascertain the likelihood that open reading frames are indeed protein-coding. As with the other measures, the CAI should be useful in this context, particularly in locating genes of moderate to high expression. However, some of the points outlined above indicate that difficulties may

arise in interpreting low CAI values. Thus, while a high CAI is probably a good indication that a reading frame is protein-coding, a low CAI may indicate a gene of low expression, a gene of heterologous origin (as with the hemolysin genes), or a noncoding region that happens to contain no termination codons. The CAI value expected for a random sequence can easily be calculated, but a relatively high value for a noncoding sequence may arise simply because DNA is not a random sequence of nucleotides, or because there is a coding sequence on the complementary strand (31). For example, an E.coli gene with no UUA, CUA or UCA codons, but otherwise having the typical codon composition of a nonhighly expressed gene (6), would give rise to an in phase open reading frame on the complementary strand with a CAI of approximately 0.28, which is similar to the lower values seen for E.coli genes (Figure 1) and somewhat higher than the value (about 0.17) expected for a random sequence.

#### Comparison of CAI with other indices.

The CAI is a very simple measure of the extent of synonymous codon usage bias, specifically in the direction of the bias seen in highly expressed genes. It has the advantage, compared with indices which measure only the frequency of certain optimal codons, of taking account of all 59 codons where synonymous alternatives exist, each in a quantitative manner. For example, both the codon bias index (4) and the frequency of optimal codons (1) treat GCU and GCC equally, as preferred codons for Ala in yeast, and yet the frequency of GCU is approximately three times that of GCC in very highly expressed genes (Table 1). With heterologous gene expression in mind it may be of primary importance to know the frequency of particularly disadvantageous codons in a gene. Simpler indices compound these very rare codons with others not in the 'optimal' category. Thus in E.coli AUA and AUU are treated equally (1), despite their very different frequency of use (see Table 1, and Ref.6). Again the CAI takes account of these differences quantitatively.

The codon preference statistic (10) is similar but not identical to the CAI<sub>obs</sub> given by equation [4]. One difference is that in calculating the codon preference statistic the p values (analogous to RSCU in equation [4]) are adjusted to take account of base composition. Another difference is that the CAI value is scaled to allow for the different amino acid compositions of different proteins (see equation [3]), and has a range from 0 - 1.0. Although this scaling cannot completely compensate for differing amino acid compositions, it facilitates comparisons between genes.

Our discussion of the use of the Codon Adaptation Index has focussed on unicellular organisms because the determinants of codon usage in multicellular organisms are not well understood (1). For example, it appears that the mammalian genome comprises regions of quite different G+C content (32), and that local G+C content is an important influence on codon usage in any one gene (1). Also tRNA abundancies are important selective constraints on codon usage, and in multicellular organisms tRNA populations vary among tissues. We also note that the only mammalian ribosomal protein genes for which DNA sequence data are available (two from mouse and two from rat -- see Ref.33) do not seem to show particularly high synonymous codon bias. It may be possible in the near future to derive a reference set of RSCU values from other highly expressed mammalian genes, and/or it may prove necessary to take into account the tissue in which the gene is expressed, for example by having several reference sets.

### ACKNOWLEDGMENTS

We thank R. Grantham for suggestions and Robert J. Schwartz for plotting the figure. This study was supported by NIH grant GM 30998.

\*To whom correspondence should be addressed

### REFERENCES

1. Ikemura, T. (1985) *Mol. Biol. Evol.* 2, 13-34.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nucleic Acids Res.* 9, r43-r79.
3. Gouy, M. and Gautier, C. (1982) *Nucleic Acids Res.* 10, 7055-7074.
4. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3026-3031.
5. Sharp, P.M., Tuohy, T.M.F. and Mosurski, K.R. (1986) *Nucleic Acids Res.* 14, 5125-5143.
6. Sharp, P.M. and Li, W-H. (1986) *Nucleic Acids Res.* 14, 7737-7749.
7. McLachlan, A.D., Staden, R. and Boswell, D.R. (1984) *Nucleic Acids Res.* 12, 9567-9575.
8. Ikemura, T. (1981) *J. Mol. Biol.* 146, 1-21.
9. Ikemura, T. (1982) *J. Mol. Biol.* 158, 573-598.
10. Gribskov, M., Devereux, J. and Burgess, R.R. (1984) *Nucleic Acids Res.* 12, 539-549.
11. Burton, Z.F., Gross, C.A., Watanabe, K.K. and Burgess, R.R. (1983) *Cell* 32, 335-349.
12. Foster, D.L. and Fillingame, R.H. (1982) *J. Biol. Chem.* 257, 2009-2013.
13. Holm, L. (1986) *Nucleic Acids Res.* 14, 3075-3087.
14. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) *Nucleic Acids Res.* 12, 6663-6671.
15. Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F. (1985) *Nucleic Acids Res.* 13, 4113-4123.
16. Pedersen, S. (1984) *EMBO J.* 3, 2895-2898.
17. Varenne, S., Buc, J., Lloubes, R. and Lazdunski, C. (1984) *J. Mol. Biol.* 180, 549-576.

18. Varenne, S. and Lazdunski, C. (1986) *J. Theor. Biol.* 120, 99-110.
19. Sharp, P.M., Rogers, M.S. and McConnell, D.J. (1985) *J. Mol. Evol.* 21, 150-160.
20. Grantham, R., Greenland, T., Louail, S., Mouchiroud, D., Prato, J.L., Gouy, M. and Gautier, C. (1985) *Bull. Inst. Pasteur* 83, 95-148.
21. Toh, H. (1986) *FEBS Letters* 194, 245-248.
22. Sharp, P.M. (1986) *Mol. Biol. Evol.* 3, 75-83.
23. Felmlee, T., Pellett, S. and Welch, R.A. (1985) *J. Bact.* 163, 94-105.
24. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
25. Li, W-H., Luo, C-C and Wu, C-I. (1985) in *Molecular Evolutionary Genetics*, MacIntyre, R.J. Ed., pp. 1-94, Plenum Press, New York.
26. Li, W-H., Gojobori, T. and Nei, M. (1981) *Nature* 292, 237-239.
27. Miyata, T. and Hayashida, H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5739-5743.
28. Sharp, P.M. and Li, W-H. (1986) *J. Mol. Evol.* (in press)
29. Li, W-H. (1987) *J. Mol. Evol.* (in press)
30. Ochman, H. and Selander, R.K. (1984) *Proc. Natl. Acad. Sci. USA* 81, 198-201.
31. Sharp, P.M. (1985) *Nucleic Acids Res.* 13, 1389-1397.
32. Bernardi, G., Olofsson, B., Filipinski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) *Science* 228, 953-958.
33. GenBank, Genetic Sequence Data Bank, Release 42.0 (1986) Bolt, Beranek and Newman.



---

**Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity**

---

Paul M.Sharp\*, Elizabeth Cowe, Desmond G.Higgins, Denis C.Shields, Kenneth H.Wolfe and Frank Wright

---

Department of Genetics, Trinity College, Dublin 2, Ireland

---

Received May 21, 1988; Revised and Accepted July 26, 1988

---

**ABSTRACT**

The genetic code is degenerate, but alternative synonymous codons are generally not used with equal frequency. Since the pioneering work of Grantham's group (1,2) it has been apparent that genes from one species often share similarities in codon frequency; under the "genome hypothesis" (1,2) there is a species-specific pattern to codon usage.

However, it has become clear that in most species there are also considerable differences among genes (3-7). Multivariate analyses have revealed that in each species so far examined there is a single major trend in codon usage among genes, usually from highly biased to more nearly even usage of synonymous codons. Thus, to represent the codon usage pattern of an organism it is not sufficient to sum over all genes (8), as this conceals the underlying heterogeneity. Rather, it is necessary to describe the trend among genes seen in that species. We illustrate these trends for six species where codon usage has been examined in detail, by presenting the pooled codon usage for the 10% of genes at either end of the major trend (Table 1).

Closely-related organisms have similar patterns of codon usage, and so the six species in Table 1 are representative of wider groups. For example, with respect to codon usage, *Salmonella typhimurium* closely resembles *E.coli* (9), while all mammalian species so far examined (principally mouse, rat and cow) largely resemble humans (4,8).

**CAUSES OF WITHIN-SPECIES DIVERSITY**

Biased codon usage may result from a combination of several factors, viz. biases in the pattern of mutation, (translational) selection among synonymous codons, or selection against particular structures in DNA. Within-species heterogeneity in codon usage has been most clearly elucidated in *E.coli*; the major trend is from a strong bias towards a particular subset of codons in highly expressed genes to more even codon usage in lowly expressed genes (3,4,7). The heavily favoured codons in highly expressed *E.coli* genes are those best recognised by the most abundant tRNA species (3,4), and it seems clear that selection mediated by the translation process can occur among alternative synonymous codons (10,11). In contrast, most of the deviation from equal synonym use in the lowly expressed genes is likely to reflect nonrandom patterns of mutation (7,12). Then the pattern of bias in a particular gene reflects a mutation-selection balance at a point determined by the strength of translational selection on that gene (7,9,12).

Similar observations have been made for *S.cerevisiae* (4,5,12,13). In *B.subtilis* (14) and *S.pombe* (15) there are similar trends among genes, but there is less information about tRNA abundances. The pattern of codon

Table 1. Codon usage diversity within six species.

		<u>E.coli</u>		<u>B.subtilis</u>		<u>S.cerevisiae</u>		<u>S.pombe</u>		<u>Drosophila</u>		<u>Human</u>		
		high	low	high	low	high	low	high	low	high	low	G+C	A+T	
Phe	UUU	0.34	1.33	0.70	1.48	0.19	1.38	0.44	1.28	0.12	0.86	0.27	1.20	UUU
	UUC	1.66	0.67	1.30	0.52	1.81	0.62	1.56	0.72	1.88	1.14	1.73	0.80	UUC
Leu	UUA	0.06	1.24	2.71	0.66	0.49	1.49	0.28	1.79	0.03	0.62	0.05	0.99	UUA
	UUG	0.07	0.87	0.00	1.03	5.34	1.48	2.16	0.80	0.69	1.05	0.31	1.01	UUG
Leu	CUU	0.13	0.72	2.13	1.24	0.02	0.73	2.44	1.55	0.25	0.80	0.20	1.26	CUU
	CUC	0.17	0.65	0.00	0.93	0.00	0.51	1.13	0.31	0.72	0.90	1.42	0.80	CUC
	CUA	0.04	0.31	1.16	0.34	0.15	0.95	0.00	0.87	0.06	0.60	0.15	0.57	CUA
	CUG	5.54	2.20	0.00	1.80	0.02	0.84	0.00	0.68	4.25	2.04	3.88	1.38	CUG
Ile	AUU	0.48	1.38	0.91	1.38	1.26	1.29	1.53	1.77	0.74	1.27	0.45	1.60	AUU
	AUC	2.51	1.12	1.96	1.14	1.74	0.66	1.47	0.59	2.26	0.95	2.43	0.76	AUC
	AUA	0.01	0.50	0.13	0.48	0.00	1.05	0.00	0.64	0.00	0.78	0.12	0.64	AUA
Met	AUG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	AUG
Val	GUU	2.41	1.09	1.88	0.83	2.07	1.13	1.61	2.04	0.56	0.74	0.09	1.32	GUU
	GUC	0.08	0.99	0.25	1.49	1.91	0.76	2.39	0.65	1.59	0.93	1.03	0.69	GUC
	GUA	1.12	0.63	1.38	0.76	0.00	1.18	0.00	1.06	0.06	0.53	0.11	0.80	GUA
	GUG	0.40	1.29	0.50	0.92	0.02	0.93	0.00	0.24	1.79	1.80	2.78	1.19	GUG
Ser	UCU	2.81	0.78	3.45	0.77	3.26	1.56	3.14	1.33	0.87	0.55	0.45	1.63	UCU
	UCC	2.07	0.60	0.00	0.81	2.42	0.81	2.57	0.52	2.74	1.41	2.09	0.80	UCC
	UCA	0.06	0.95	1.50	1.29	0.08	1.30	0.00	1.56	0.04	0.84	0.26	1.23	UCA
	UCG	0.00	1.04	0.00	0.94	0.02	0.66	0.00	0.67	1.17	1.30	0.68	0.13	UCG
Pro	CCU	0.15	0.75	2.29	0.99	0.21	1.17	2.00	1.21	0.42	0.43	0.58	1.50	CCU
	CCC	0.02	0.68	0.00	0.27	0.02	0.75	2.00	0.83	2.73	1.02	2.02	0.83	CCC
	CCA	0.42	1.03	1.14	1.08	3.77	1.38	0.00	1.51	0.62	1.04	0.36	1.57	CCA
	CCG	3.41	1.54	0.57	1.66	0.00	0.70	0.00	0.45	0.23	1.51	1.04	0.10	CCG
Thr	ACU	1.87	0.76	2.21	0.39	1.83	1.23	1.89	1.52	0.65	0.70	0.36	1.45	ACU
	ACC	1.91	1.29	0.00	0.98	2.15	0.78	2.11	1.04	3.04	1.58	2.37	0.92	ACC
	ACA	0.10	0.68	1.38	1.64	0.00	1.38	0.00	1.04	0.10	0.77	0.36	1.45	ACA
	ACG	0.12	1.28	0.41	0.98	0.01	0.60	0.00	0.40	0.21	0.95	0.92	0.18	ACG
Ala	GCU	2.02	0.61	2.94	0.78	3.09	1.07	2.30	1.79	0.95	0.91	0.45	1.59	GCU
	GCC	0.18	1.18	0.08	1.14	0.89	0.76	1.49	0.50	2.82	1.93	2.38	0.92	GCC
	GCA	1.09	0.79	0.60	1.19	0.03	1.49	0.21	1.14	0.09	0.59	0.36	1.38	GCA
	GCG	0.71	1.42	0.38	0.89	0.00	0.68	0.00	0.57	0.14	0.57	0.82	0.11	GCG

Relative Synonymous Codon Usage (RSCU; Ref.5) values are presented for two groups of genes from each of six species: Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens.

(An RSCU value is the observed number of codons divided by the number expected if all codons for that amino acid were used equally.)

Table 1 (cont.)

		<u>E.coli</u>		<u>B.subtilis</u>		<u>S.cerevisiae</u>		<u>S.pombe</u>		<u>Drosophila</u>		<u>Human</u>	
		high	low	high	low	high	low	high	low	high	low	G+C	A+T
Tyr	UAU	0.38	1.28	0.50	1.29	0.06	1.13	0.48	1.24	0.23	0.96	0.34	1.17
	UAC	1.63	0.72	1.50	0.71	1.94	0.87	1.52	0.76	1.77	1.04	1.66	0.83
	UAA	--	--	--	--	--	--	--	--	--	--	--	--
	UAG	--	--	--	--	--	--	--	--	--	--	--	--
His	CAU	0.45	1.21	2.00	1.28	0.32	1.16	0.56	1.44	0.29	0.86	0.30	1.28
	CAC	1.55	0.79	0.00	0.72	1.68	0.84	1.44	0.56	1.71	1.14	1.70	0.72
Gln	CAA	0.12	0.76	1.71	0.88	1.98	1.10	1.85	1.67	0.03	0.88	0.21	0.98
	CAG	1.88	1.24	0.29	1.13	0.02	0.90	0.15	0.33	1.97	1.12	1.79	1.02
Asn	AAU	0.02	1.12	0.47	1.21	0.06	1.28	0.30	1.41	0.13	1.13	0.33	1.20
	AAC	1.98	0.88	1.53	0.79	1.94	0.72	1.70	0.59	1.87	0.87	1.67	0.80
Lys	AAA	1.63	1.50	1.83	1.47	0.16	1.24	0.10	1.27	0.06	0.81	0.34	1.17
	AAG	0.37	0.50	0.17	0.53	1.84	0.76	1.90	0.73	1.94	1.19	1.66	0.83
Asp	GAU	0.51	1.43	0.53	1.16	0.70	1.38	0.78	1.56	0.90	1.10	0.36	1.29
	GAC	1.49	0.57	1.47	0.84	1.30	0.62	1.22	0.44	1.10	0.90	1.64	0.71
Glu	GAA	1.64	1.28	1.40	1.27	1.98	1.29	0.69	1.20	0.19	0.73	0.26	1.33
	GAG	0.36	0.72	0.60	0.73	0.02	0.71	1.31	0.80	1.81	1.27	1.74	0.67
Cys	UGU	0.60	0.94	0.00	0.94	1.80	1.10	0.14	1.56	0.07	0.71	0.42	1.09
	UGC	1.40	1.06	2.00	1.06	0.20	0.90	1.86	0.44	1.93	1.29	1.58	0.91
ter	UGA	--	--	--	--	--	--	--	--	--	--	--	--
Trp	UGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Arg	CGU	4.47	1.71	3.11	0.54	0.63	0.64	5.17	1.89	2.65	0.69	0.38	0.64
	CGC	1.53	2.41	1.78	1.21	0.00	0.39	0.83	0.26	3.07	1.55	2.72	0.36
	CGA	0.00	0.52	0.00	0.74	0.00	0.65	0.00	0.86	0.07	1.12	0.31	0.81
	CGG	0.00	0.80	0.00	0.81	0.00	0.34	0.00	0.43	0.00	1.12	1.53	0.51
Ser	AGU	0.13	1.01	0.45	0.56	0.06	0.97	0.14	1.48	0.04	0.89	0.31	1.26
	AGC	0.93	1.62	0.60	1.63	0.16	0.70	0.14	0.44	1.13	1.01	2.22	0.94
Arg	AGA	0.00	0.37	1.11	2.02	5.37	2.51	0.00	1.71	0.00	0.56	0.22	2.40
	AGG	0.00	0.19	0.00	0.67	0.00	1.47	0.00	0.86	0.21	0.95	0.84	1.28
Gly	GGU	2.27	1.29	1.38	0.54	3.92	1.32	3.36	1.87	1.34	0.91	0.34	0.84
	GGC	1.68	1.31	0.97	1.30	0.06	0.92	0.59	0.27	1.66	1.65	2.32	0.76
	GGA	0.00	0.64	1.66	1.24	0.00	1.22	0.05	1.60	0.99	0.98	0.29	1.79
	GGG	0.04	0.76	0.00	0.92	0.02	0.55	0.00	0.27	0.00	0.46	1.05	0.61

For each species, genes have been ranked according to their position along the major intraspecific trend in codon bias (see text). The highest 10% and the lowest 10% of genes have been drawn from: 165 E.coli genes (7), 76 B.subtilis genes (8,14), 154 S.cerevisiae genes (5,8), 40 S.pombe genes (15), 84 D.melanogaster genes (16) and 290 human genes (8). The sample size for S.pombe is rather small, but the codon frequencies appear to be reliable (15). Full gene listings are available from the authors.

frequencies in lowly expressed genes from *B.subtilis* is most strongly indicative of mutational bias (14).

Recently, we have reported evidence of selection among synonymous codons in the multicellular organism *D.melanogaster* (16). In contrast, among human genes the major variation is in G+C content associated with the local base composition around the gene (6). This variation has not been attributed to translational selection, and is most easily explained in terms of variation in mutation biases among chromosomal regions.

### CODON BIAS RANKINGS

For *E.coli*, *B.subtilis*, *S.cerevisiae* and *S.pombe* codon bias in a gene is measured by the Codon Adaptation Index (CAI). A species-specific reference set of very highly expressed genes is used to assess the relative fitness of each synonymous codon, and the CAI for a gene is then calculated as the geometric mean of the fitness values for each codon in that gene. (For a full description, see Ref.17.)

Since the biological basis of codon frequencies in *Drosophila* is not yet so firmly established (for example, there may be more than one optimal set of codons, depending on the tissue of gene expression) we have simply estimated codon bias as the deviation from equal synonym use, by a "chi-square" scaled by gene length (16); this index is very highly correlated with the major trend among genes. Finally, human genes are ranked by G+C content at silent positions, since this is the major source of variation among genes (4,6).

FORTTRAN 77 programs to calculate these indices are available (on IBM-type floppy disks) from the authors on request.

**ACKNOWLEDGEMENT:** This is a publication from the Irish National Centre for Bioinformatics. This work was supported in part by a grant from the European Community Biotechnology Action Programme.

\*To whom correspondence should be addressed

### REFERENCES

1. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) *Nucleic Acids Res.* 8, r49-r62.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nucleic Acids Res.* 9, r43-r74.
3. Gouy, M. and Gautier, C. (1982) *Nucleic Acids Res.* 10, 7055-7074.
4. Ikemura, T. (1985) *Mol. Biol. Evol.* 2, 13-34.
5. Sharp, P.M., Tuohy, T.M.F. and Mosurski, K.R. (1986) *Nucleic Acids Res.* 14, 5125-5143.
6. Aota, S. and Ikemura, T. (1986) *Nucleic Acids Res.* 14, 6345-6355.
7. Sharp, P.M. and Li, W-H. (1986) *Nucleic Acids Res.* 14, 7737-7749.
8. Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) *Nucleic Acids Res.* 16, r315-r402.
9. Ikemura, T. (1985) in *Molecular Evolution and Population Genetics*, Aoki, K. and Ohta, T. Eds., pp. 385-406, Springer-Verlag, Berlin.

10. de Boer, H. and Kastelein, R.A. (1986) in From gene to protein; steps dictating the maximal level of gene expression, Reznikoff, W.S. and Gold, L. Eds., pp. 225-283, Butterworths, Mass.
11. Kurland, C.G. (1987) Trends Biochem. Sci. 12, 126-128.
12. Bulmer, M. (1988) J. Evol. Biol. 1, 15-26.
13. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
14. Shields, D.C. and Sharp, P.M. (1987) Nucleic Acids Res. 15, 8023-8040.
15. Sharp, P.M. and Wright, F. (1988) in preparation
16. Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F. (1988) Mol. Biol. Evol. 5 (in press)
17. Sharp, P.M. and Li, W-H. (1987) Nucleic Acids Res. 15, 1281-1295.

## A Complex Mutant of TEM-1 $\beta$ -Lactamase with Mutations Encountered in Both IRT-4 and Extended-Spectrum TEM-15, Produced by an *Escherichia coli* Clinical Isolate

D. SIROT,<sup>1\*</sup> C. RECULE,<sup>2</sup> E. B. CHAIBI,<sup>3</sup> L. BRET,<sup>1</sup> J. CROIZE,<sup>2</sup> C. CHANAL-CLARIS,<sup>1</sup>  
R. LABIA,<sup>3</sup> AND J. SIROT<sup>1</sup>

Laboratoire de Bactériologie-Virologie, Faculté de Médecine, 63001 Clermont-Ferrand Cedex,<sup>1</sup> Service de Bactériologie et Virologie, Centre Hospitalier Universitaire de Grenoble, 38043 Grenoble Cedex 9,<sup>2</sup> and UMR 175 Centre National de la Recherche Scientifique-MNHN, 29000 Quimper,<sup>3</sup> France

Received 28 October 1996/Returned for modification 3 February 1997/Accepted 6 April 1997

*Escherichia coli* GR102 was isolated from feces of a leukemic patient. It expressed different levels of resistance to amoxicillin or ticarcillin plus clavulanate and to the various cephalosporins tested. The double-disk synergy test was weakly positive. Production of a  $\beta$ -lactamase with a pI of 5.6 was transferred to *E. coli* HB101 by conjugation. The nucleotide sequence was determined by direct sequencing of the amplification products obtained by PCR performed with TEM gene primers. This enzyme differed from TEM-1 (*bla*T-1B gene) by four amino acid substitutions: Met→Leu-69, Glu→Lys-104, Gly→Ser-238 and Asn→Asp-276. The amino acid substitutions Leu-69 and Asp-276 are known to be responsible for inhibitor resistance of the IRT-4 mutant, as are Lys-104 and Ser-238 substitutions for hydrolytic activity of the extended-spectrum  $\beta$ -lactamases TEM-15, TEM-4, and TEM-3. These combined mutations led to a mutant enzyme which conferred a level of resistance to coamoxiclav (MIC, 64  $\mu$ g/ml) much lower than that conferred by IRT-4 (MIC, 2,048  $\mu$ g/ml) but higher than that conferred by TEM-15 or TEM-1 (MIC, 16  $\mu$ g/ml). In addition, the MIC of ceftazidime for *E. coli* transconjugant GR202 (1  $\mu$ g/ml) was lower than that for *E. coli* TEM-15 (16  $\mu$ g/ml) and higher than that for *E. coli* IRT-4 or TEM-1 (0.06  $\mu$ g/ml). The MICs observed for this TEM-type enzyme were related to the kinetic constants  $K_m$  and  $k_{cat}$  and the 50% inhibitory concentration, which were intermediate between those observed for IRT-4 and TEM-15. In conclusion, this new type of complex mutant derived from TEM-1 (CMT-1) is able to confer resistance at a very low level to inhibitors and at a low level to extended-spectrum cephalosporins. CMT-1 received the designation TEM-50.

Overproduction of *Escherichia coli* chromosomal  $\beta$ -lactamase is one cause of resistance to  $\beta$ -lactam- $\beta$ -lactamase inhibitor combinations such as amoxicillin (AMX)-clavulanate (CA) and also results in reduced susceptibility to all  $\beta$ -lactams except carbapenems (15).

In *E. coli*, resistance to all  $\beta$ -lactams except cephamycins and carbapenems may be caused by extended-spectrum  $\beta$ -lactamases. These enzymes are susceptible to  $\beta$ -lactamase inhibitors such as CA (10, 14, 15) and are therefore detected by synergy tests (10), and strains producing such mutants are often susceptible to  $\beta$ -lactam- $\beta$ -lactamase inhibitor combinations.

In *E. coli* isolates, the most recently discovered mechanism of resistance to AMX-CA is production of inhibitor-resistant TEM  $\beta$ -lactamases (IRT) (8).

*E. coli* GR102, isolated from feces of a leukemic patient in the hematology unit of the teaching hospital of Grenoble, France, harbored an unusual  $\beta$ -lactam resistance phenotype with resistance to AMX and ticarcillin (TIC) alone and combined with CA and resistance to all cephalosporins, including cephamycins, at various levels. In addition, the double-disk synergy test used for extended-spectrum  $\beta$ -lactamase detection was weakly positive.

This complex phenotype suggested that the  $\beta$ -lactam resistance of the strain was due to the presence of several  $\beta$ -lactamases or a combination of different mechanisms of resistance to  $\beta$ -lactams.

### MATERIALS AND METHODS

**Strains.** The strains used included *E. coli* GR102, a clinical isolate producing a novel  $\beta$ -lactamase; *E. coli* HB101, used as a recipient strain for transfer; *E. coli* HB101/p111 (TEM-1 producing); *E. coli* CF042 (IRT-4-TEM-35 producing) (8); and *E. coli* transformant DH5 $\alpha$  (CF244), obtained by electroporation from *Klebsiella pneumoniae* Kp240 (TEM-15 producing) (16).

**Susceptibility to  $\beta$ -lactams.** The MICs of AMX, TIC, cephalothin (CF), cefotaxime (CTX), ceftazidime (CAZ), aztreonam (ATM), cefepime (FEP), and cefpirome (CPO) alone and combined with CA at a fixed concentration of 2  $\mu$ g/ml were determined. A method of dilution with Mueller-Hinton agar (Sanofi Diagnostics Pasteur, Marnes-la-Coquette, France) and an inoculum of  $10^4$  CFU per spot were used. Antibiotics were provided as powders by SmithKline Beecham Pharmaceuticals (AMX, TIC, and CA), Roussel-Uclaf (CTX and CPO), Glaxo Wellcome Research and Development (CAZ), and Bristol-Myers Squibb (ATM and FEP).

Detection of extended-spectrum  $\beta$ -lactamase was performed with the double-disk synergy test as described by Jarlier et al. (10).

**Isoelectric focusing.** Isoelectric focusing was performed with polyacrylamide gels containing ampholines with a pH range of 3.5 to 10.0 as previously described (19), and  $\beta$ -lactamases with known pIs (TEM-1 [pI 5.4], TEM-2 [pI 5.6], TEM-15 [pI 6.0], and IRT-4 [pI 5.2]) were used as standards.

**Transfer experiment.** A transfer experiment was performed with *E. coli* GR102 and the recipient *E. coli* HB101. Transconjugants were selected on agar containing rifampin (300  $\mu$ g/ml) and gentamicin (8  $\mu$ g/ml) or CAZ (0.5  $\mu$ g/ml).

**Sequencing of DNA amplified by PCR.** On the assumption that the transconjugant strain contained *bla*<sub>TEM</sub>, a single-stranded DNA template was generated for sequencing by PCR performed with an asymmetric ratio of amplification primers A and B, and the nucleotide sequence was determined as previously described (3), by direct sequencing of the amplified product obtained from the transconjugant *E. coli* GR202.

**Determination of  $\beta$ -lactamase kinetic parameters  $k_{cat}$ ,  $K_m$ , and  $k_{cat}/K_m$ .** Affinity ( $K_m$ ) and catalytic activity ( $k_{cat}$ ) were determined with highly purified extracts ( $\geq 97\%$  pure) by using a computerized microacidimetric method (13).

\* Corresponding author. Mailing address: Laboratoire de Bactériologie-Virologie, Faculté de Médecine, 28 Place Henri-Dunant, 63001 Clermont-Ferrand Cedex, France. Phone: 33 4 73 60 80 18. Fax: 33 4 73 27 74 94.

TABLE 1. Nucleotide and amino acid substitutions in *bla*<sub>TEM</sub> genes

Nucleotide no. <sup>a</sup>	Nucleotide (amino acid) <sup>b</sup> in:			
	<i>bla</i> <sub>TEM-1(Tn2)</sub>	<i>bla</i> <sub>TEM-15</sub>	<i>bla</i> <sub>IRT-4</sub>	<i>bla</i> <sub>CMT-1</sub>
226	T (Phe)	C	T	T
317	C (Gln-39)	C	C	C
346	A (Glu)	A	G	A
407	A (Met-69)	A	C (Leu)	C (Leu)
436	T (Gly)	C	T	T
512	G (Glu-104)	A (Lys)	G	A (Lys)
604	T (Ala)	G	G	T
682	T (Thr)	T	T	T
914	G (Gly-238)	A (Ser)	G	A (Ser)
925	G (Gly)	G	G	G
1022	A (Asn-276)	A	G (Asp)	G (Asp)

<sup>a</sup> Nucleotide numbering is according to Sutcliffe (21).<sup>b</sup> The amino acid is indicated when a point mutation leads to an amino acid substitution compared with the sequences of TEM-1(Tn2). Numbering is according to Ambler et al. (1).

All  $\beta$ -lactamases were purified from crude extracts by size exclusion chromatography on Sephadex G-100 (Pharmacia), preparative isoelectric focusing, and reverse-phase high-performance liquid chromatography on a C<sub>18</sub> Nucleosil 500A column (Interchim) as described by Brun et al. (2). The homogeneity of the preparations was determined by analytical sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The  $k_{cat}$ ,  $K_m$ , and  $k_{cat}/K_m$  values of enzyme CMT-1 were compared with those of the  $\beta$ -lactamases TEM-1, TEM-15, and IRT-4. The kinetics of TEM-1 and its mutants toward penicillins and cephalosporins were compared. Inhibition studies of TEM-1 and the mutant enzymes with CA, sulbactam, and tazobactam were performed. The affinity of the enzyme for the inhibitor, expressed as the inhibition constant ( $K_i$ ), was measured by using competition procedures with benzylpenicillin. It is determined from the extrapolated rate at the time when the inhibitor is added. The 50% inhibitory concentration ( $IC_{50}$ ) was determined after incubation of the inhibitor and the enzyme for 10 min (completed inactivation) at 37°C before measurement of the remaining enzymatic activity. The  $IC_{50}$  is defined as the inhibitor concentration causing 50% inhibition of benzylpenicillin hydrolysis by the enzyme.

## RESULTS

**Resistance phenotype of *E. coli* GR102.** *E. coli* GR102 expressed a complex  $\beta$ -lactam resistance phenotype with resistance to AMX and TIC alone and combined with CA and resistance to cephalosporins at various levels: high-level resistance to narrow-spectrum cephalosporins and low-level resistance to extended-spectrum cephalosporins (MICs, 1 to 32  $\mu$ g/ml). A positive synergy test with CA suggested the presence of a mutant extended-spectrum  $\beta$ -lactamase of class A origin. In addition, this strain had reduced susceptibility to cefoxitin (MIC, 128  $\mu$ g/ml) and, to a lesser extent, cefotetan and moxa-

lactam (data not shown). This reduced susceptibility to cephalosporins was probably related to decreased permeability of the strain for  $\beta$ -lactams, since *E. coli* GR102 had lost an outer membrane protein with a molecular mass of 40 kDa (data not shown). Overproduction of the chromosomal cephalosporinase was not detected.

This strain was also resistant to aminoglycosides (tobramycin, gentamicin, and netilmicin), probably owing to production of an AAC (3)-II enzyme, and to chloramphenicol, tetracyclines, and sulfonamides.

**Conjugative transfer and isoelectric focusing.** The gene encoding resistance to  $\beta$ -lactams, except cephalosporins, was transferred by conjugation from *E. coli* GR102 to rifampin-resistant *E. coli* HB101 (GR202). Selection for CAZ or gentamicin resistance revealed the transfer of an 85-kb plasmid conferring resistance to  $\beta$ -lactams, aminoglycosides (tobramycin, gentamicin, and netilmicin), tetracyclines, and sulfonamides.

By isoelectric focusing, two bands at pIs 5.4 and 5.6 were observed in *E. coli* GR102 and one band at pI 5.6 was observed in *E. coli* transconjugant GR202.

**Nucleotide sequencing.** As shown in Table 1, from the transconjugant GR202 producing a  $\beta$ -lactamase with a pI of 5.6, nucleotide sequencing revealed a *bla*<sub>TEM</sub> gene identical to the *bla*<sub>T-IB</sub> gene (Tn-2) at positions 226, 317, 346, 436, 604, 682, and 925, which discriminate the *bla*<sub>TEM</sub> genes (4, 7).

The *bla*<sub>TEM</sub> gene from the *E. coli* transconjugant differed from the *bla*<sub>T-IB</sub> gene by four point mutations. These mutations consisted of the nucleotide change A→C at position 407, which leads to the amino acid substitution Met→Leu at position 69 (1); the nucleotide change G→A at positions 512 and 914, leading to the amino acid substitutions Glu→Lys at position 104 and Gly→Ser at position 238; and the nucleotide change A→G at position 1022, leading to the amino acid substitution Asn→Asp at position 276. The two amino acid substitutions at positions 69 and 276 are observed in the IRT-4-TEM-35 enzyme (2, 8, 22), and the two amino acid substitutions at positions 104 and 238 are observed in the extended-spectrum  $\beta$ -lactamase TEM-15 (16).

**$\beta$ -Lactam MICs for TEM mutants.** Consequently, we compared the MICs of  $\beta$ -lactams for *E. coli* GR102 and its transconjugant GR202, producing a complex mutant form of TEM (CMT-1), with MICs for IRT-4-producing *E. coli* CF0042, TEM-15-producing *E. coli* CF244, and TEM-1-producing *E. coli* HB101 (Table 2).

For *E. coli* GR202 producing CMT-1, the MICs of AMX-CA (64  $\mu$ g/ml) and TIC-CA (64  $\mu$ g/ml) were much lower than

TABLE 2. MICs of  $\beta$ -lactams for *E. coli* CMT-1 (GR102 and its transconjugant, GR202), *E. coli* TEM-15 (CF244), *E. coli* IRT-4 (CF0042), and *E. coli* TEM-1 (HB101)

<i>E. coli</i> strain (enzyme)	MIC ( $\mu$ g/ml) of:															
	AMX		TIC		CF		CTX		CAZ		ATM		FEP		CPO	
	Alone	With CA <sup>a</sup>	Alone	With CA	Alone	With CA	Alone	With CA	Alone	With CA	Alone	With CA	Alone	With CA	Alone	With CA
GR102 (CMT-1)	4,096	256	>4,096	512	256	128	4	0.5	4	1	1	0.25	8	1	32	4
GR202 <sup>b</sup> (CMT-1)	2,048	64	4,096	64	8	4	1	0.03	1	0.12	0.12	0.06	1	0.03	2	0.12
CF244 <sup>c</sup> (TEM-15)	>4,096	16	>4,096	32	128	4	8	0.06	16	0.25	4	0.12	1	0.03	2	0.03
CF0042 (IRT-4)	4,096	2,048	1,024	512	8	4	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06
HB101 (TEM-1)	4,096	16	4,096	32	8	2	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06
HB101 <sup>d</sup>	4	4	1	1	4	4	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06	≤0.06

<sup>a</sup> CA was used at 2  $\mu$ g/ml.<sup>b</sup> *E. coli* HB101 transconjugant.<sup>c</sup> *E. coli* DH5 $\alpha$  transformant.<sup>d</sup> *E. coli* recipient strain.

TABLE 3. Production of TEM-type  $\beta$ -lactamases in *E. coli*

Enzyme	pI	Producing organism	Activity in crude extract <sup>a</sup>	Sp act <sup>b</sup> of purified protein
TEM-1	5.4	HB101	2.2	2.48
TEM-35-IRT-4	5.2	CF0042	2.1	2.17
TEM-15	6.0	CF244	0.4	0.083
TEM-50-CMT-1	5.6	GR202	0.4	0.23

<sup>a</sup> Micromoles of benzylpenicillin per minute per milligram of protein.<sup>b</sup> Micromoles of benzylpenicillin per minute per microgram of protein. Determined with highly purified preparations ( $\geq 97\%$  pure).

those observed for IRT-4-producing strain CF0042 (2,048 and 512  $\mu\text{g/ml}$ , respectively). Similarly, for strain GR202, MICs of CTX (1  $\mu\text{g/ml}$ ) and CAZ (1  $\mu\text{g/ml}$ ) were lower than those observed for TEM-15-producing strain CF244 (8 and 16  $\mu\text{g/ml}$ , respectively) and higher than those for IRT-4 producing strain CF0042 ( $\leq 0.06$   $\mu\text{g/ml}$ ). The same 1:8 ratio of MICs was observed for aztreonam (0.12  $\mu\text{g/ml}$  for the CMT-1 producer and 4  $\mu\text{g/ml}$  for the TEM-15 producer). The MICs of cefepime and cefpirome (1 and 2  $\mu\text{g/ml}$ ) were identical for the CMT-1 and TEM-15 producers.

*E. coli* GR202 (CMT-1) was 2 to 4 times less susceptible to AMX or TIC plus CA (64  $\mu\text{g/ml}$ ) and 16 times less susceptible to CTX, CAZ, FEP, and CPO (1 to 2  $\mu\text{g/ml}$ ) than was *E. coli* HB101 (TEM-1<sup>+</sup>). MICs of  $\beta$ -lactam substrates in the presence of 2 and 4  $\mu\text{g}$  of sulbactam or tazobactam per ml were about fourfold lower for the CMT-1-producing strain than for the TEM-1-producing strain (data not shown).

**Enzymatic and kinetic parameters of  $\beta$ -lactamases.** Enzymatic and kinetic parameters of the new complex mutant enzyme CMT-1 with regard to penicillins and cephalosporins were compared with those of the TEM-1, IRT-4, and TEM-15  $\beta$ -lactamases (Tables 3 to 5). The specific activity of the highly purified CMT-1 protein was 10-fold lower than that of TEM-1 (Table 3).

For all penicillins, the  $k_{\text{cat}}$  values of CMT-1 were about 10-fold lower than those of TEM-1 and IRT-4 and about twice as high as those of TEM-15. The catalytic efficiencies ( $k_{\text{cat}}/K_m$ ) of the three mutant enzymes were lower than those of the

TABLE 4. Comparison of the kinetics<sup>a</sup> of TEM-1 and its mutant forms

Drug	$k_{\text{cat}}$ (s <sup>-1</sup> ), $K_m$ ( $\mu\text{M}$ ), $k_{\text{cat}}/K_m$ ( $\mu\text{M}^{-1}\text{s}^{-1}$ )			
	TEM-1	TEM-35-IRT-4	TEM-15	TEM-50-CMT-1
Benzylpenicillin	1,200, 25, 48.0	1,050, 140, 7.5	40, 6, 6.7	110, 17, 6.5
AMX	920, 26, 35.4	900, 245, 8.5	26, 5, 5.2	70, 33, 2.1
TIC	115, 10, 11.5	125, 320, 0.4	8, 2, 4.0	15, 30, 0.5
Carbenicillin	132, 13, 10.2	120, 360, 0.3	7, 3, 2.3	13, 60, 0.2
Piperacillin	987, 45, 21.9	945, 320, 2.9	64, 12, 5.3	111, 31, 3.6
CF	122, 250, 0.5	52, 1,200, 0.04	43, 23, 1.7	62, 324, 0.2
Cephaloridine	2,045, 800, 2.5	340, 1,420, 0.2	37, 30, 1.2	320, 310, 1.03
Cefoperazone	470, 260, 1.8	305, 1,325, 0.2	25, 22, 1.1	150, 118, 1.3
Cefuroxime	ND <sup>b</sup>	ND	24, 91, 0.3	20, 260, 0.08
Ceftriaxone	ND	ND	92, 50, 1.8	35, 385, 0.09
CTX	1.2, ND	ND	180, 100, 1.8	150, 873, 0.2
CAZ	ND	ND	7, 80, 0.1	3, ND
ATM	ND	ND	0.2, ND	1, ND

<sup>a</sup> The standard deviation for analysis was  $\leq 10\%$ .<sup>b</sup> ND, not detected; the rate was too small to determine  $k_{\text{cat}}$  and  $K_m$  reliably.

TABLE 5. Inhibition of TEM-1 and its mutant forms by CA, sulbactam, and tazobactam

Enzyme	$\text{IC}_{50}$ ( $\mu\text{M}$ ), $K_i$ ( $\mu\text{M}$ )		
	CA	Sulbactam	Tazobactam
TEM-1	0.08, 0.1	6.1, 0.9	0.1, 0.01
TEM-35-IRT-4	28, 27	304, 49	1.8, 0.6
TEM-15	0.01, 0.02	0.03, 0.02	0.01, 0.008
TEM-50-CMT-1	0.25, 0.7	0.5, 0.4	0.04, 0.06

TEM-1 enzyme, and the values observed for CMT-1 and IRT-4 with carboxy- and ureidopenicillins were similar.

For cephalosporins,  $k_{\text{cat}}$  values of CMT-1 were slightly lower than or similar to those of TEM-15 for ceftriaxone, CTX, CAZ, and ATM; however, the catalytic efficiencies of the CMT-1 enzyme were only 5 to 11% of those of TEM-15 with ceftriaxone and CTX. No activity of TEM-1 or IRT-4 against expanded-spectrum cephalosporins and ATM ( $k_{\text{cat}}$ s of  $< 1 \text{ s}^{-1}$  associated with  $K_s$  of  $\geq 500 \mu\text{M}$ ) was detected.

The  $\text{IC}_{50}$  of CA for CMT-1 (Table 5) was higher (0.25  $\mu\text{M}$ ) than that for TEM-1 (0.08  $\mu\text{M}$ ) and TEM-15 (0.01  $\mu\text{M}$ ) but 100-fold lower than that for IRT-4 (28  $\mu\text{M}$ ). Sulbactam was the least efficient inhibitor of IRT-4 ( $\text{IC}_{50}$ , 304  $\mu\text{M}$ ), while its inhibitor efficiency was similar to that of CA for CMT-1 ( $\text{IC}_{50}$ , 0.5  $\mu\text{M}$ ). Tazobactam was the most efficient inhibitor of all of these  $\beta$ -lactamases. Moreover, CMT-1 ( $\text{IC}_{50}$ , 0.04  $\mu\text{M}$ ) and TEM-15 ( $\text{IC}_{50}$ , 0.01  $\mu\text{M}$ ) were more susceptible to inhibition by tazobactam than was TEM-1 ( $\text{IC}_{50}$ , 0.1  $\mu\text{M}$ ).

## DISCUSSION

The TEM-1 derivative described in this report constitutes a new type of complex mutant, CMT-1, combining mutations responsible for inhibitor resistance (Leu-69 and Asp-276) and those responsible for extended-spectrum activity (Lys-104 and Ser-238). It is the first example of such a  $\beta$ -lactamase produced by a clinical isolate of *E. coli*.

**Mutations conferring resistance to  $\beta$ -lactam inhibitors.** Replacement of methionine 69, just adjacent to serine 70, by aliphatic amino acids such as leucine influences the positioning of residues (5, 17) because the buried side chain at position 69 lies behind  $\beta$ -strand B3, forming the back wall of the oxyanion pocket in which the  $\beta$ -lactam's carbonyl group is polarized (12). Moreover, crystallographic data indicate that residues in the C-terminal  $\alpha$  helix, such as Asn-276, restrict the mobility of the Arg-244 side chain and so play a role in maintaining the integrity of the active site (11). Because small  $\beta$ -lactams such as CA must rely primarily on attractive interactions with the oxyanion hole and Arg-244, inhibitor resistance exists in the natural variant IRT-4, containing changes at residues 69 and 276 (12). This IRT-4 mutant enzyme is one of the most resistant to inhibition by CA among the IRT-type enzymes (2, 22), with a CA  $\text{IC}_{50}$  350-fold higher than that for TEM-1. The CA resistance of this mutant is confirmed by high-level resistance to combinations of AMX and TIC with CA (MIC, 2,048 and 512  $\mu\text{g/ml}$ , respectively). Inhibition studies showed that the CMT-1 enzyme was 100-fold less resistant than the IRT-4 mutant but only 3 times as resistant to inhibition by CA as the wild-type TEM-1  $\beta$ -lactamase. These kinetic results were closely related to the moderate resistance level (64  $\mu\text{g/ml}$ ) of the CMT-1-producing strain to AMX-CA or TIC-CA. The inhibitor resistance usually caused by the Leu-69 mutation may be decreased by the close proximity of the Ser-238 mutation (see below).



**Mutations conferring extended-spectrum activity.** The Glu→Lys change at position 104 contributes to the precise positioning of residues 130 to 132 (SDN loop), which are involved in substrate binding, but seems insufficient alone to confer true resistance to expanded-spectrum cephalosporins (18, 20).

It is generally recognized that the substitution Gly→Ser-238 enlarges the active site, thereby creating an enzyme with increased affinity for the 7-oxymino cephalosporins (6). All of the TEM variants reported to have Ser-238 contain methionine at position 69, and mutant enzyme CMT-1 is the first harboring both mutations Ser-238 and Leu-69. The affinity may be affected by a change at position 69, since the side chain at position 238, on the inner side of the B3  $\beta$ -strand, lies very close to the side chain of residue 69 (12).

The TEM-1 variant, with the associated changes Glu→Lys-104 and Gly→Ser-238, is TEM-15 (16). Complex mutant CMT-1, with these last mutations, conferred a lower level of resistance to CTX and CAZ than did TEM-15, and this difference correlated with the kinetic constants. The kinetic comparison of substrate hydrolysis in extended-spectrum  $\beta$ -lactamases TEM-15 and CMT-1 revealed that the catalytic efficiency ( $k_{cat}/K_m$ ) of CMT-1 was lower than that of TEM-15 for CTX (10-fold) and ceftriaxone (20-fold).

Overall, the hydrolytic properties of this complex TEM mutant enzyme were found to be closer to those of an extended-spectrum enzyme than to those of an inhibitor-resistant enzyme. However, the predominant effect of the mutations Lys-104 and Ser-238, which are responsible for extended-spectrum activity and inhibitor hypersusceptibility, was clearly attenuated by the mutations Leu-69 and Asp-276. In the strain producing CMT-1, complete reversal of CA resistance by mutations enhancing activity against 7-oxymino cephalosporins was not observed as reported in a Ser-164-Ser-244 mutant obtained by site-specific mutagenesis (9).

*E. coli* GR102 was isolated from a patient treated with CPO (4 g/day) and amikacin (1 g/day) for 12 days. An *E. coli* strain with a typical  $\beta$ -lactam inhibitor resistance phenotype and the same resistances to other antibiotics had been isolated from the same sample (feces) from this patient a week before. This suggests that an inhibitor-resistant TEM mutant enzyme with the Leu-69 and Asp-276 mutations emerged first, and then, under antibiotic (CPO) and mutagenic agent (cytarabine and daunorubicin) pressure, this mutant underwent the two additional mutations, Lys-104 and Ser-238, responsible for extended-spectrum activity. Unfortunately, this hypothesis could not be confirmed since the initial *E. coli* IRT-producing strain was not kept.

In conclusion, the production of this complex TEM mutant cannot alone account for the high-level multiresistance to  $\beta$ -lactams of the *E. coli* GR102 isolate, in which several resistance mechanisms were involved (TEM-1 and, probably, decreased permeability). It would probably be more beneficial for an *E. coli* strain to produce two different TEM mutants (an extended-spectrum mutant and an inhibitor-resistant mutant) simultaneously than to produce one double mutant. If each mutant conferred its own resistance phenotype, high-level resistance to both CA combinations and extended-spectrum cephalosporins could then be expected.

#### ACKNOWLEDGMENTS

We thank J. M. Frère for the gift of highly purified TEM-15 extract and P. Courvalin for the gift of *K. pneumoniae* Kp240 producing the TEM-15 enzyme. We thank Rolande Perroux, Marlène Jan, and Dominique Rubio for technical assistance.

This work was supported in part by a grant from the Direction de la Recherche et des Etudes doctorales, Ministère de l'Éducation Nationale, Paris, France.

#### REFERENCES

1. Ambler, R. P., A. F. N. Coulson, J. M. Frère, J. M. Ghuyssen, B. Joris, M. Forsman, R. C. Levesque, G. Tiraby, and S. G. Waley. 1991. A standard numbering scheme for the class A  $\beta$ -lactamases. *Biochem. J.* 276:269–272.
2. Brun, T., J. Péduzzi, M. M. Canica, G. Paul, P. Névoit, M. Barthélémy, and R. Labia. 1994. Characterization and amino acid sequence of IRT-4, a novel TEM-type enzyme with a decreased susceptibility to  $\beta$ -lactamase inhibitors. *FEMS Microbiol. Lett.* 120:111–118.
3. Chanal, C., M. C. Poupard, D. Sirot, R. Labia, J. Sirot, and R. Cluzel. 1992. Nucleotide sequences of CAZ-2, CAZ-6, and CAZ-7  $\beta$ -lactamase genes. *Antimicrob. Agents Chemother.* 36:1817–1820.
4. Chen, S. T., and R. C. Clowes. 1987. Variations between the nucleotide sequences of Tn1, Tn2, and Tn3 and expression of  $\beta$ -lactamase in *Pseudomonas aeruginosa* and *Escherichia coli*. *J. Bacteriol.* 169:913–916.
5. Delaire, M., R. Labia, J. P. Samama, and J. M. Masson. 1992. Site-directed mutagenesis at the active site of *Escherichia coli* TEM-1  $\beta$ -lactamase. *J. Biol. Chem.* 267:20600–20606.
6. Du Bois, S. K., M. S. Marriott, and S. G. B. Amyes. 1995. TEM- and SHV-derived extended-spectrum  $\beta$ -lactamases: relationship between selection, structure and function. *J. Antimicrob. Chemother.* 35:7–22.
7. Goussard, S., and P. Courvalin. 1991. Sequence of the genes *blaT-1B* and *blaT-2*. *Gene* 102:71–73.
8. Henquell, C., C. Chanal, D. Sirot, R. Labia, and J. Sirot. 1995. Molecular characterization of nine different types of mutants among 107 inhibitor-resistant TEM  $\beta$ -lactamases from clinical isolates of *Escherichia coli*. *Antimicrob. Agents Chemother.* 39:427–430.
9. Imtiaz, U., E. K. Manavathu, S. Mobashery, and S. A. Lerner. 1994. Reversal of clavulanate resistance conferred by a Ser-244 mutant of TEM-1  $\beta$ -lactamase as a result of a second mutation (Arg to Ser at position 164) that enhances activity against ceftazidime. *Antimicrob. Agents Chemother.* 38:1134–1139.
10. Jarlier, V., M. H. Nicolas, G. Fournier, and A. Philippon. 1988. Extended broad-spectrum  $\beta$ -lactamases conferring transferable resistance to newer  $\beta$ -lactam agents in *Enterobacteriaceae*: hospital prevalence and susceptibility patterns. *Rev. Infect. Dis.* 10:867–878.
11. Jelsch, C., F. Lenfant, J. M. Masson, and J. P. Samama. 1992.  $\beta$ -Lactamase TEM-1 of *E. coli*. Crystal structure determination at 2.5 Å resolution. *FEBS Lett.* 299:135–142.
12. Knox, J. R. 1995. Extended-spectrum and inhibitor-resistant TEM-type  $\beta$ -lactamases: mutations, specificity, and three-dimensional structure. *Antimicrob. Agents Chemother.* 39:2593–2601.
13. Labia, R., J. Andriillon, and F. Le Goffic. 1973. Computerized microacidimetric determination of  $\beta$ -lactamase Michaelis-Menten constants. *FEBS Lett.* 33:42–44.
14. Labia, R., A. Morand, K. Tiwari, J. S. Pitton, D. Sirot, and J. Sirot. 1988. Kinetic properties of two plasmid-mediated  $\beta$ -lactamases from *Klebsiella pneumoniae* with strong activity against third-generation cephalosporins. *J. Antimicrob. Chemother.* 21:301–307.
15. Livermore, D. M. 1995.  $\beta$ -Lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* 8:557–584.
16. Mabilat, C., and P. Courvalin. 1990. Development of "oligotyping" for characterization and molecular epidemiology of TEM  $\beta$ -lactamases in members of the family *Enterobacteriaceae*. *Antimicrob. Agents Chemother.* 34:2210–2216.
17. Oliphant, A. R., and K. Struhl. 1989. An efficient method for generating proteins with enzymatic properties: application to  $\beta$ -lactamase. *Proc. Natl. Acad. Sci. USA* 86:9094–9098.
18. Petit, A., L. Maveyraud, F. Lenfant, J. P. Samama, R. Labia, and J. M. Masson. 1995. Multiple substitutions at position 104 of  $\beta$ -lactamase TEM-1: assessing the role of this residue in substrate specificity. *Biochem. J.* 305:33–40.
19. Sirot, D., C. Chanal, C. Henquell, R. Labia, J. Sirot, and R. Cluzel. 1994. Clinical isolates of *Escherichia coli* producing multiple TEM-mutants resistant to  $\beta$ -lactamase inhibitors. *J. Antimicrob. Chemother.* 33:1117–1126.
20. Soweik, J. A., B. Singer, S. Ohringer, M. F. Mally, T. J. Dougherty, J. Z. Gougoutas, and K. Bush. 1991. Substitution of lysine at position 104 or 240 of TEM  $\beta$ -lactamase enhances the effect of serine-164 substitution on hydrolysis or affinity for cephalosporins and the monobactam aztreonam. *Biochemistry* 30:3179–3188.
21. Sutcliffe, G. 1978. Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. *Proc. Natl. Acad. Sci. USA* 75:3737–3741.
22. Zhou, X. Y., F. Bordon, D. Sirot, M. D. Kitzis, and L. Gutmann. 1994. Emergence of clinical isolates of *Escherichia coli* producing TEM-1 derivatives or an OXA-1  $\beta$ -lactamase conferring resistance to  $\beta$ -lactamase inhibitors. *Antimicrob. Agents Chemother.* 38:1085–1089.

## Structure-Function Relationships among Wild-Type Variants of *Staphylococcus aureus* $\beta$ -Lactamase: Importance of Amino Acids 128 and 216

RAMA KISHAN R. VOLADRI,<sup>1\*</sup> MURALI K. R. TUMMURU,<sup>1</sup> AND DOUGLAS S. KERNODLE<sup>1,2</sup>

Division of Infectious Diseases, Department of Medicine, Vanderbilt University, Nashville, Tennessee 37232-2605,<sup>1</sup>  
and Department of Veterans Affairs Medical Center, Nashville, Tennessee 37232<sup>2</sup>

Received 30 May 1996/Accepted 10 October 1996

$\beta$ -Lactamases inactivate penicillin and cephalosporin antibiotics by hydrolysis of the  $\beta$ -lactam ring and are an important mechanism of resistance for many bacterial pathogens. Four wild-type variants of *Staphylococcus aureus*  $\beta$ -lactamase, designated A, B, C, and D, have been identified. Although distinguishable kinetically, they differ in primary structure by only a few amino acids. Using the reported sequences of the A, C, and D enzymes along with crystallographic data about the structure of the type A enzyme to identify amino acid differences located close to the active site, we hypothesized that these differences might explain the kinetic heterogeneity of the wild-type  $\beta$ -lactamases. To test this hypothesis, genes encoding the type A, C, and D  $\beta$ -lactamases were modified by site-directed mutagenesis, yielding mutant enzymes with single amino acid substitutions. The substitution of asparagine for serine at residue 216 of type A  $\beta$ -lactamase resulted in a kinetic profile indistinguishable from that of type C  $\beta$ -lactamase, whereas the substitution of serine for asparagine at the same site in the type C enzyme produced a kinetic type A mutant. Similar bidirectional substitutions identified the threonine-to-alanine difference at residue 128 as being responsible for the kinetic differences between the type A and D enzymes. Neither residue 216 nor 128 has previously been shown to be kinetically important among serine-active-site  $\beta$ -lactamases.

$\beta$ -Lactam antibiotics, including the penicillins and cephalosporins, are important agents in the therapy of bacterial infections. However, in some clinical settings the usefulness of these agents has been diminished by the emergence and spread of bacterial strains that produce  $\beta$ -lactamase, which hydrolyzes the  $\beta$ -lactam ring and inactivates the drug's antimicrobial effect (27). This problem has been demonstrated most dramatically with *Staphylococcus aureus*. Whereas the vast majority of clinical isolates of *S. aureus* were highly susceptible to penicillin G at the time of its introduction into clinical use in the early 1940s, the spread of  $\beta$ -lactamase-producing, penicillin-resistant strains was so widespread by the late 1940s that penicillin G was no longer a reliable antistaphylococcal agent. Most clinical isolates of *S. aureus* produce  $\beta$ -lactamase (36).

Four types of *S. aureus*  $\beta$ -lactamase have been identified by serologic (34, 35) and kinetic (19, 20) methods. These variants originally were designated types A, B, C, and D by Richmond (34) and Rosdahl (35). This nomenclature should not be confused with that of the different classes of  $\beta$ -lactamases, A through D, that has been used more recently to group the  $\beta$ -lactamases of all bacterial species on the basis of active site (serine versus zinc), size, and kinetic characteristics (4). Each of the four recognized types of *S. aureus*  $\beta$ -lactamase (A, B, C, and D) is a class A  $\beta$ -lactamase with a serine active site. The mature form of the enzyme has a molecular mass of 30 kDa, contains 257 amino acids, and is excreted extracellularly (1).

The type A and C staphylococcal  $\beta$ -lactamases are easily distinguished kinetically, either by substrate profile (18) or by

$K_m$  and catalytic constant ( $k_{cat}$ ) determinations (43). Whereas the kinetics of hydrolysis exhibited by the type A and D staphylococcal  $\beta$ -lactamases are similar (20), they can be rapidly and reproducibly distinguished by a fivefold difference in the ratio of the rates of the initial velocity of hydrolysis of penicillin G and cefazolin (43). The lower penicillin G/cefazolin hydrolysis ratio of the type D  $\beta$ -lactamase appears to be related to a lower  $k_{cat}$  for the hydrolysis of penicillin G by purified type D enzyme than the other staphylococcal  $\beta$ -lactamases (43).

The genes encoding the type A, C, and D  $\beta$ -lactamases of *S. aureus* are located on plasmids and have been cloned and sequenced (5, 8, 9, 40). The deduced amino acid sequences identify six amino acid differences in the primary sequence of the prototypic type A  $\beta$ -lactamase from strains PC1 and S1 compared with the type C sequence from strain 3804, including amino acids 116, 202, 205, 206, 212, and 216. In addition, there are five differences, at amino acids 93, 121, 128, 226, and 229, between the prototypic type A  $\beta$ -lactamase and the type D  $\beta$ -lactamase produced by strain FAR4. The type C and D enzymes differ by 11 amino acids.

The molecular basis for the kinetic heterogeneity exhibited by the staphylococcal  $\beta$ -lactamases has not been determined. However, the structure of the type A variant of *S. aureus*  $\beta$ -lactamase is known from X-ray crystallographic analysis (11, 12), and several of the amino acids which are different between the type A, C, and D  $\beta$ -lactamases are located close to the active-site cleft. To test the hypothesis that single amino acid differences at sites close to the active site are responsible for the kinetic heterogeneity exhibited by naturally occurring variants of *S. aureus*  $\beta$ -lactamase, we constructed mutant  $\beta$ -lactamases with single amino acid substitutions at sites where the wild-type enzymes differ and evaluated the kinetics of the mutant enzyme.

\* Corresponding author. Mailing address: Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232. Phone: (615) 327-4751, ext. 5486. Fax: (615) 343-6160. Electronic mail address: voladrrr@ctrvax.vanderbilt.edu.

TABLE 1. Plasmids and phagemids used in this study

Plasmid	Host	Resistance <sup>a</sup>	Description	Source or reference
pS1	<i>S. aureus</i>	Amp	From strain S1, carries <i>blaZ</i> encoding type A $\beta$ -lactamase	18
pII3804	<i>S. aureus</i>	Amp	From strain 3804, carries <i>blaZ</i> encoding type C $\beta$ -lactamase	9
pUB101	<i>S. aureus</i>	Amp	From strain FAR4, carries <i>blaZ</i> encoding type D $\beta$ -lactamase	25
pE194	<i>S. aureus</i>	Em	From strain RN2442, carries gene encoding Em <sup>r</sup>	13
pBC SK+	<i>E. coli</i>	Cm	Phagemid vector	Stratagene
pTZ18R	<i>E. coli</i>	Amp	Phagemid vector	Bio-Rad
pVK100	<i>S. aureus-E. coli</i>	Em, Cm	<i>E. coli-S. aureus</i> shuttle plasmid constructed by cloning 3.7-kb pE194 on a <i>SacI</i> site into <i>Sall</i> of pBC SK+	This study
pVK101	<i>S. aureus-E. coli</i>	Em, Cm	pVK100 with a 4-kb <i>EcoRI-HindIII</i> fragment carrying $\beta$ -lactamase regulatory genes from p13796 but lacks <i>blaZ</i>	This study
pVK102	<i>E. coli</i>	Amp	pTZ18R with 4-kb <i>HincII</i> fragment with type A <i>blaZ</i> from pS1	This study
pVK103	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.1-kb <i>HindIII</i> fragment with type A <i>blaZ</i> from pVK102	This study
pVK104	<i>E. coli</i>	Amp	pBC SK+ with 8-kb <i>EcoRI</i> fragment with type C <i>blaZ</i> from pII3804	This study
pVK105	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.8-kb <i>HindIII</i> fragment with type C <i>blaZ</i> from pVK104	This study
pVK106	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.3-kb <i>HindIII</i> fragment with type D <i>blaZ</i> from pUB101	This study
pVK107	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.1-kb <i>HindIII</i> type A <i>blaZ</i> fragment except a single nucleotide substitution leading to S216N change	This study
pVK108	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.1-kb <i>HindIII</i> type A <i>blaZ</i> fragment except a single nucleotide substitution leading to T128A change	This study
pVK109	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.8-kb <i>HindIII</i> type C <i>blaZ</i> fragment except a single nucleotide substitution leading to N216S change	This study
pVK110	<i>E. coli-S. aureus</i>	Amp, Em	pVK101 with 1.3-kb <i>HindIII</i> type D <i>blaZ</i> fragment except a single nucleotide substitution leading to A128T change	This study

<sup>a</sup> Amp, ampicillin; Cm, chloramphenicol; Em, erythromycin.

## MATERIALS AND METHODS

**Chemicals and media.** Standard powders of nitrocefin (BBL Microbiology Systems, Cockeysville, Md.); cephaloridine (Sigma Chemical Co., St. Louis, Mo.); methicillin, ampicillin, and cephalirin (Bristol Laboratories, Syracuse, N.Y.); and cefazolin and penicillin G (Eli Lilly & Co., Indianapolis, Ind.) were used to prepare antibiotic solutions for kinetic studies. Cation-exchange resin P11 (Whatman Laboratories, Kent, England) and *meta*-aminophenyl boronic acid hemisulfate and succinamide-activated sepharose (Sigma) were used for preparing columns for  $\beta$ -lactamase purification. Restriction endonucleases, T4 DNA ligase, and Sequenase were purchased from United States Biochemical Corp., Cleveland, Ohio. A Muta Gene Phagemid kit (Bio-Rad Laboratories, Fullerton, Calif.) was used for site-directed mutagenesis. The oligonucleotides required for site-directed mutagenesis and sequencing of the  $\beta$ -lactamase gene were synthesized on a Cyclone Plus Automatic DNA synthesizer (Millipore, Bedford, Mass.) by the DNA Core Facility, Department of Molecular Physiology and Biophysics, Vanderbilt University. Modified 1% CY broth was prepared as described by Novick (29). LB and 2 $\times$  YT media were prepared according to the methods of Maniatis et al. (26). Tryptic soy broth was purchased from BBL.

**Bacterial strains, plasmids, and cultivation.** The plasmids and phagemids used and constructed during this study are listed in Table 1. *S. aureus* RN4220 (R. P. Novick, Public Health Research Institute, New York, N.Y.) was used as a recipient for protoplast transformation. *Escherichia coli* DH5 $\alpha$  was used for the transformation and propagation of *E. coli* plasmids and *S. aureus-E. coli* shuttle plasmids except when indicated otherwise.

**Construction of *S. aureus*  $\beta$ -lactamase expression vector.** An *E. coli-S. aureus* shuttle vector, pVK101, was constructed with erythromycin as a selectable marker in *S. aureus*, by using a strategy similar to that employed by East et al. (8) to construct *S. aureus*  $\beta$ -lactamase expression vector pAE706. First, *S. aureus* plasmid pE194 was digested with *Sall* and cloned into the *SacI* site of pBC SK+ (Stratagene, La Jolla, Calif.) to create pVK100. Next *blaR* ( $\beta$ -lactamase regulatory gene) from *S. aureus* plasmid p13796 was cloned on a 4-kb *HindIII-EcoRI* fragment into pVK100 to produce pVK101.

**Cloning of type A, C, and D *blaZ*.** Large-scale isolation of plasmid DNA from *S. aureus* by ultracentrifugation in a cesium chloride gradient (Var lac old Chemical Co., Inc., Bergenfield, N.J.) was performed by methods described by Galetto et al. (10). A 4-kb *HincII* fragment carrying the *blaZ* ( $\beta$ -lactamase structural gene) from pS1 was cloned into pTZ18R to create pVK102. *HindIII* digestion of pVK102 enabled the type A *blaZ* to be mobilized on a 1.1-kb fragment which was cloned into pVK101. This produced pVK103 in which *blaZ* and *blaR* are transcribed divergently from the same intracistronic region. A similar strategy was used to clone the type C *blaZ* from pII3804 into pVK101 via a pBC SK+ intermediate (pVK104) to produce pVK105. The plasmid DNA from pUB101 was digested with *HindIII*, and a 1.3-kb fragment carrying type D *blaZ* was cloned in the desired orientation into pVK101 to produce pVK106.

**Site-directed mutagenesis of  $\beta$ -lactamase.** Oligonucleotide-directed mutagenesis was performed by the method of Kunkel et al. (23, 24). *HindIII* fragments

containing the type A, C, and D *blaZ* genes were cloned individually into the phage vector M13mp18 (42). Two oligonucleotides, each with a single mismatch (underlined), were used to introduce specific mutations at amino acids 128 (5'-TCACTATATGCGCATTTGAAGCC3'; Thr to Ala) and 216 (5'-AGTATCTC CGTTTTTATTATT3'; Ser to Asn) of the type A *blaZ*. Additional primers were constructed to produce the reverse mutation in the type C and D *blaZ* genes (i.e., 5'-AGTGTCTCCGCTTTTATTATT, Asn to Ser, type C *blaZ* and 5'-ATCACT ATATGTCATTGAAGC, Ala to Thr, type D *blaZ*). Single-stranded DNA from the six randomly picked plaques was sequenced by the dideoxy chain termination method (38) using Sequenase enzyme to confirm the desired mutation. The complete open reading frame was sequenced in order to rule out the presence of any spurious mutations.

**Expression of wild-type and mutant  $\beta$ -lactamases in *S. aureus* RN4220.** Wild-type *blaZ* encoding type A, C, and D  $\beta$ -lactamases and mutant *blaZ* genes from M13mp18 recombinants were cloned as 1.1- to 1.8-kb *HindIII* fragments into the *HindIII* site of *S. aureus-E. coli* shuttle plasmid pVK101. Recombinants were selected on LB medium containing ampicillin (100  $\mu$ g/ml) and chloramphenicol (30  $\mu$ g/ml). The desired orientation of the insert was verified either by restriction with *EcoRV*, which has a single site within type A and D *blaZ* and one just upstream of the *blaZ* promoter within *blaR*, or by PCR. Protoplast transformation was used to transfer the recombinant shuttle plasmids into *S. aureus* RN4220 (6).

**$\beta$ -Lactamase purification from *S. aureus*.** The wild-type  $\beta$ -lactamase-producing strains and the RN4220 transformants were grown in 5 liters of modified 1% CY medium containing methicillin (0.5  $\mu$ g/ml) or 2-(2'-carboxyphenyl)benzoyl-6-aminopenicillanic acid (7.5  $\mu$ M, to induce  $\beta$ -lactamase production) at 37°C and 150 rpm for 18 h. The extracellular  $\beta$ -lactamases were purified by sequential cation-exchange and affinity chromatography (21).

**Enzyme kinetics.** Initial velocities of hydrolysis were monitored at a wavelength corresponding to the maximal change in absorbance between the unhydrolyzed substrate and the hydrolyzed product, which included the following: cephaloridine, 254 nm; cefazolin, 272 nm; nitrocefin, 482 nm; cephalirin, 258 nm; ampicillin, 235 nm; and penicillin G, 232 nm (37, 39).  $\beta$ -Lactamase assays were performed in 0.1 M sodium phosphate buffer, pH 6.0, in 1-cm cuvettes at 37°C with a DU-70 recording spectrophotometer (Beckman Instruments, Fullerton, Calif.). For  $K_m$  and  $k_{cat}$  determinations, assays of the initial velocity of hydrolysis were performed using 100, 50, 33.3, 20, 14.2, and 11.1  $\mu$ M solutions of each cephalosporin antibiotic. Penicillin hydrolysis assays were performed at initial substrate concentrations of 1,000, 500, 333, 200, 142, and 111  $\mu$ M. The maximal rate of hydrolysis,  $V_{max}$ , and the Michaelis constant,  $K_m$ , for each substrate-enzyme combination were determined from  $[s]/v$ -against- $[s]$  plots (where  $[s]$  is substrate concentration and  $v$  is velocity) (41) with computerized software (Hyper; Department of Biochemistry, University of Liverpool, Liverpool, United Kingdom). The turnover number,  $k_{cat}$ , was calculated from the  $V_{max}$  by using a molecular mass of the purified  $\beta$ -lactamase of 30,000 g/mol. Mean values and standard error of the mean values were calculated from the results of

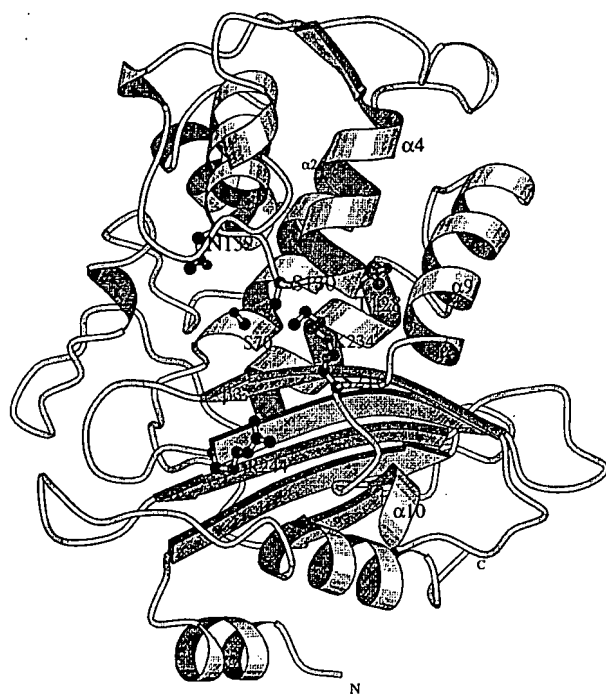


FIG. 1. Tertiary structure of *S. aureus* PC1  $\beta$ -lactamase. The active-site cleft of the  $\beta$ -lactamase is located at the left side of the  $\beta$ 3 strand. The active-site amino acids are S-70, S-130, N-132, K-234, and R-244. Amino acids T-128 and S-216 were substituted by site-directed mutagenesis of the cloned structural gene. The diagram was generated by the computer software MOLSCRIPT (22) using X-ray crystallographic coordinates stored in the Brookhaven Protein Database (accession no. 3BLM).

8 to 12  $[s/v]$ -against- $[s]$  plots for each enzyme-substrate combination by using computer software (Minitab Data Analysis Software, release 10.2; Minitab, Inc., State College, Pa.).

## RESULTS

**Construction and expression of mutant  $\beta$ -lactamases.** The primary goal of this study was to determine the functional domains, particularly the specific amino acid(s), responsible for the kinetic heterogeneity observed among naturally occurring variants of *S. aureus*  $\beta$ -lactamase. Mutant  $\beta$ -lactamases with single amino acid substitutions at residues 128 (Thr to

Ala) and 216 (Ser to Asn) were constructed by using oligonucleotide-directed mutagenesis and the type A  $\beta$ -lactamase gene (*blaZ*) as a template. In addition, the reverse mutations were introduced by using the type D (Ala to Thr, residue 128) and type C (Asn to Ser, residue 216) *blaZ* genes as template DNA. These residues were selected primarily because of the amino acids which differed among the type A, C, and D  $\beta$ -lactamases, 128 and 216 were closest to the active-site cleft (Fig. 1). Each *blaZ* was cloned into *E. coli*-*S. aureus* shuttle plasmid pVK101, which was then transformed into *S. aureus* RN4220.  $\beta$ -Lactamase production was induced, and the wild-type  $\beta$ -lactamases expressed by the reference strains and the mutant enzymes expressed in RN4220 were purified to homogeneity.

**Effect of substitution at amino acid 216 on the kinetics of type A and C  $\beta$ -lactamases.**  $\beta$ -Lactams which have been shown previously to be useful in distinguishing between the wild-type *S. aureus*  $\beta$ -lactamases (20) were used to compare the kinetics of hydrolysis of the reference and mutant enzymes (Tables 2 and 3). Between the type A and type C  $\beta$ -lactamases there were a 10-fold difference in the  $K_m$  values of cefazolin, a 5-fold difference in the  $K_m$  values of cephalirin, and a 5-fold difference in the  $k_{cat}$  values of nitrocefin. These differences appear to be due to whether Ser or Asn was present at residue 216. Replacement of Ser by Asn in the type A  $\beta$ -lactamase yielded a mutant (A, S216N) that was closer kinetically to the type C than the type A enzyme (e.g., cefazolin  $K_m$  values: 167  $\mu$ M, mutant; 145  $\mu$ M, type C; 13.1  $\mu$ M, type A). The reverse mutation using type C *blaZ* DNA for site-directed mutagenesis yielded a mutant (C, N216S) that was similar to the type A enzyme (cefazolin  $K_m$  value, 15.3  $\mu$ M). Also, the differences between the kinetic type A and type C *S. aureus*  $\beta$ -lactamases are clearly shown by comparing relative efficiency of hydrolysis values (Table 4). The  $k_{cat}$  values of most substrates other than nitrocefin did not clearly distinguish between the type A and type C enzymes (Table 3).

**Effect of substitution at amino acid 128 on the kinetics of type A and D  $\beta$ -lactamases.** The reference type A  $\beta$ -lactamase had a cefazolin  $K_m$  value that was threefold lower than and an ampicillin  $K_m$  value that was twofold greater than the respective  $K_m$  values of the type D enzyme. In addition, the  $k_{cat}$  values of ampicillin and penicillin G were three- to fourfold higher with the type A compared to the type D  $\beta$ -lactamase. These differences were related to whether Ala or Thr was present at residue 128. Replacement of Thr by Ala in the type A  $\beta$ -lactamase yielded a mutant (A, T128A) that was closer kinetically to the type D than the type A enzyme (e.g., penicillin  $k_{cat}$  values: 47  $s^{-1}$ , mutant; 254  $s^{-1}$ , type A; 66  $s^{-1}$ , type

TABLE 2.  $K_m$  values of  $\beta$ -lactam antibiotics for purified wild-type and mutant  $\beta$ -lactamases of *S. aureus*<sup>a</sup>

Antibiotic	Mean $K_m$ [ $\mu$ M (SEM)] for $\beta$ -lactamase						
	Wild type			Altered by site-directed mutagenesis			
	pS1, type A	pH3804, type C	pUB101, type D	pVK107 (A, S216N)	pVK108 (A, T128A)	pVK109 (C, N216S)	pVK110 (D, $\Delta$ 128T)
Cephaloridine	4.3 (1.0)	6.3 (1.8)	4.6 (1.2)	6.5 (1.7)	5.3 (1.3)	4.5 (0.7)	4.8 (1.1)
Cefazolin	13.1 (2.8)	<b>145</b> (18.7)	<b>38.3</b> (9.3)	<b>167</b> (35)	<b>43.9</b> (6.6)	<b>15.3</b> (2.7)	<b>13.3</b> (1.8)
Cephapirin	4.5 (1.2)	<b>24.9</b> (1.7)	7.1 (1.7)	<b>27.3</b> (3.9)	7.0 (0.8)	<b>5.5</b> (1.5)	4.5 (1.0)
Nitrocefin	5.1 (1.3)	<b>10.5</b> (2.3)	7.1 (1.6)	<b>11.6</b> (2.3)	6.8 (1.5)	<b>6.7</b> (1.5)	6.6 (1.7)
Ampicillin	<b>195</b> (35)	128 (22)	<b>119</b> (22)	153 (15)	<b>118</b> (15)	208 (31)	<b>219</b> (43)
Penicillin G	29.2 (6.5)	25.5 (6.9)	26.0 (8.2)	29.0 (9.1)	34.3 (9.4)	31.3 (8.6)	19.7 (6.5)

<sup>a</sup> The initial velocities of hydrolysis of 100, 50, 33.3, 20, 14.2, and 11.1  $\mu$ M solutions of the cephalosporins and 1,000, 500, 333, 200, 142, and 111  $\mu$ M solutions of ampicillin and penicillin G were monitored with a recording spectrophotometer.  $K_m$  values were determined by the use of  $[s/v]$ -against- $[s]$  plots. Each value represents the mean from 8 to 12 determinations.  $K_m$  values that are altered by amino acid replacements are in boldface.

TABLE 3.  $k_{cat}$  values of  $\beta$ -lactam antibiotics for purified wild-type and mutant  $\beta$ -lactamases of *S. aureus*

Antibiotic	Mean $k_{cat}$ value [ $s^{-1}$ (SEM)] <sup>a</sup> for $\beta$ -lactamase						
	Wild type			Altered by site-directed mutagenesis			
	pS1, type A	pI13804, type C	pUB101, type D	pVK107 (A, S216N)	pVK108 (A, T128A)	pVK109 (C, N216S)	pVK110 (D, A128T)
Cephaloridine	1.5 (0.3)	2.6 (0.2)	0.97 (0.2)	3.4 (0.4)	0.83 (0.1)	0.74 (0.17)	1.37 (0.1)
Cefazolin	1.4 (0.1)	1.8 (0.1)	2.1 (0.1)	2.51 (0.3)	2.24 (0.1)	0.86 (0.1)	0.92 (0.1)
Cephapirin	0.37 (0.01)	0.6 (0.01)	0.21 (0.02)	0.75 (0.1)	0.19 (0.03)	0.27 (0.02)	0.36 (0.01)
Nitrocefin	33.7 (4.6)	6.9 (0.4)	16.3 (2.0)	8.0 (0.9)	15.3 (3.4)	33.1 (4.1)	37.0 (4.4)
Ampicillin	<b>560</b> (49)	249 (14)	<b>133</b> (23)	353 (28)	<b>91</b> (10)	392 (33)	<b>423</b> (46)
Penicillin G	<b>254</b> (17)	140 (5)	65.6 (6)	162 (14)	47.3 (6.8)	173 (16)	192 (12)

<sup>a</sup> Mean value in molecules of antibiotic hydrolyzed per second per molecule of enzyme, as determined from  $[s]/v$ -against- $[s]$  plots and assuming a molecular mass of the purified  $\beta$ -lactamase of 30,000 g/mol. Each value is derived from 8 to 12 determinations.  $k_{cat}$  values that are altered by amino acid replacements are in boldface.

D). The reverse mutation using type D *blaZ* DNA for site-directed mutagenesis yielded a mutant (D, A128T) that was similar to the type A enzyme (penicillin G  $k_{cat}$  value, 192  $s^{-1}$ ). The type A and D enzymes are easily distinguished by the ratio of the rates of hydrolysis of penicillin G and cefazolin (Table 5).

## DISCUSSION

Naturally occurring variants of *S. aureus*  $\beta$ -lactamase can be distinguished on the basis of the kinetics of hydrolysis of selected penicillin and cephalosporin antibiotics (20, 43). In this study we have shown that these kinetic differences are determined by single amino acid substitutions at positions close to the active site of the enzyme. Specifically, the presence of an Asn instead of Ser at residue 216 determines a type C kinetic profile, and the presence of an Ala instead of Thr at residue 128 determines a type D kinetic profile. Enzymes exhibiting the type A kinetic profile have Ser and Thr at these two sites, respectively. We also found that substitutions at some positions other than 128 and 216 where the type A, C, and D enzymes have different amino acids did not alter the kinetic profile of the enzyme (e.g., amino acid 121 [data not shown]).

This situation is reminiscent of what has been reported for the newer TEM-type  $\beta$ -lactamases and SHV  $\beta$ -lactamase variants among gram-negative bacterial species (17). Broad-spectrum TEM and SHV variant  $\beta$ -lactamases capable of hydrolyzing ceftazidime, cefotaxime, and/or other newer cephalosporins have become problematic in many medical centers in recent years, especially among isolates of *Klebsiella pneumoniae* and *E. coli* (14, 17, 31). As with the *S. aureus*

enzymes, the altered kinetic profile of the broad-spectrum TEM-type  $\beta$ -lactamases is based on modest changes in primary structure, generally one to three amino acid substitutions at sites close to the active site (32). In addition, single mutations that result in TEM and SHV variant  $\beta$ -lactamases exhibiting resistance to commercially available  $\beta$ -lactamase inhibitors such as clavulanic acid have been described (7, 15).

A major difference between the histories of the variant TEM and staphylococcal enzymes, however, is that whereas the former appear to be a consequence of selective antibiotic pressure in the clinical setting, the *S. aureus* enzymes appear to have remained remarkably stable over time. Although  $\beta$ -lactamase-producing strains of *S. aureus* spread widely during the first few years following the clinical introduction of penicillin, some clinical isolates collected and saved prior to penicillin use have been shown to produce the A and C variants of *S. aureus*  $\beta$ -lactamase. Furthermore, the prevalence of the various types of staphylococcal  $\beta$ -lactamases among clinical isolates in the United States in the 1980s (19, 20) are similar to what was reported in England by Richmond in the mid-1960s when he first described the existence of different staphylococcal  $\beta$ -lactamase serotypes (34). Despite the widespread use since the early 1960s of antistaphylococcal penicillins such as methicillin, oxacillin, and nafcillin, new staphylococcal  $\beta$ -lactamases capable of hydrolyzing these agents efficiently have not been observed.

Both of the amino acid positions that we have shown to be responsible for the kinetic differences among the wild-type *S. aureus* enzymes have not previously been cited as contributing to  $\beta$ -lactamase function. The major kinetic difference between the type A and C enzymes is the  $K_m$  values of certain cepha-

TABLE 4. Relative efficiencies of hydrolysis for  $\beta$ -lactam antibiotics of purified wild-type and mutant  $\beta$ -lactamases of *S. aureus*

Antibiotic	REH <sup>a</sup> (%) of $\beta$ -lactamase						
	Wild type			Altered by site-directed mutagenesis			
	pS1, type A	pI13804, type C	pUB101, type D	pVK107 (A, S216N)	pVK108 (A, T128A)	pVK109 (C, N216S)	pVK110 (D, A128T)
Cephaloridine	100	100	100	100	100	100	100
Cefazolin	<b>32</b>	<b>3</b>	24	<b>3</b>	28	<b>24</b>	25
Cephapirin	<b>24</b>	<b>6</b>	14	<b>5</b>	17	<b>30</b>	29
Nitrocefin	<b>1,943</b>	<b>159</b>	1,095	<b>133</b>	1,406	<b>1,980</b>	2,004
Ampicillin	<b>820</b>	<b>473</b>	562	<b>449</b>	481	<b>868</b>	690
Penicillin G	<b>2,565</b>	<b>1,336</b>	1,200	<b>1,071</b>	863	<b>2,208</b>	3,578

<sup>a</sup> Relative efficiency of hydrolysis (REH) is a relative value of  $k_{cat}/K_m$ . These values are expressed as a percentage of the REH for cephaloridine by the same  $\beta$ -lactamase (i.e., the cephaloridine REH was calculated in liters per mole per second and assigned a value of 100), determined from the mean  $k_{cat}$  and  $K_m$  values for each enzyme-substrate combination. REH values that are altered by amino acid replacements are in boldface.

TABLE 5. Ratios of the rates of hydrolysis of penicillin G and cefazolin by kinetic type A and D  $\beta$ -lactamases

$\beta$ -Lactamase (type)	Relative $k_{cat}$ ratio <sup>a</sup>	Fixed-concn ratio <sup>b</sup>
Wild type		
pS1 (A)	178.7	183.5
pUB101 (D)	32.5	36.6
Altered by site-directed mutagenesis		
pVK108 (A, T128A)	23.6	45.7
pVK110 (D, A128T)	208.5	162.5

<sup>a</sup> Ratio of the  $k_{cat}$  values of penicillin G and cefazolin.<sup>b</sup> Determined from the initial velocities of hydrolysis using a 500  $\mu$ M concentration of penicillin G and 100  $\mu$ M concentration of cefazolin.

losporins, particularly cefazolin and cephalirin (Table 2). These enzymes differ not only in hydrolysis of certain cephalosporin substrates but also in the inhibition profile of some  $\beta$ -lactamase inhibitors, including sulbactam (unpublished observations) and tazobactam (3), with the type A enzyme being more susceptible to inhibition.

The substitution at residue 216 might affect  $\beta$ -lactamase structure and function in several ways. First, the side chain of Asn is bulkier (the accessible surface area of Asn is 158 Å<sup>2</sup> [15.8 nm<sup>2</sup>] [28]) than that of Ser (122 Å<sup>2</sup> [12.2 nm<sup>2</sup>]) and may be hindering the substrate binding into the active-site cleft. Modelling studies with cefazolin docked into the PC1  $\beta$ -lactamase active-site cleft show that the side chain C $\alpha$  substituent in cefazolin is positioned close to side chain of Ser-216 (unpublished observations) and substitution of Asn for Ser at this position would result in steric hindrance. Second, the refined crystal structure of PC1  $\beta$ -lactamase at 2 Å (0.2 nm) indicated that the amino acid 216 is located on a short 3<sub>10</sub> helix comprising amino acids 215 to 217 and this helix is stabilized through a helix N capping (33) between Asn-214, which is highly conserved among class A  $\beta$ -lactamases (2), and Ser-216 (O-N, 2.9 Å [0.29 nm]). In addition to this, the side chain OH of Ser-216 is also involved in a hydrogen bond with Asn-214 side chain carbonyl (O-O, 3.3 Å [0.33 nm]). Substitution of Asn for Ser at 216 may alter the topology of this short helix. The crystal structure of a type C enzyme could help to clarify the structure of this loop, and attempts are under way to crystallize the type C enzyme. Third, amino acid 216 is located close to the  $\beta$ 3 strand, and changes at this residue might alter the relative positioning of other active-site amino acids such as the K-T-G triad.

The reason why the replacement of Thr by Ala at residue 128 should affect enzyme function is less clear. The effect of the substitution was primarily a reduction in the  $k_{cat}$  of the penicillins and nitrocefin along with modest effects on the  $K_m$  values of cefazolin and ampicillin (Table 2). The crystal structure of PC1 indicates that amino acid 128 is located at the C terminus of  $\alpha$ -helix 4 close to the active-site cleft. It is two residues away from the highly conserved S-D-N loop (amino acids 130 to 132) of the class A  $\beta$ -lactamases. The catalytic function of the S-D-N loop has been verified by site-directed mutagenesis of *Streptomyces albus* G  $\beta$ -lactamase (16) and *E. coli* TEM  $\beta$ -lactamase (30). The proximity of amino acid 128 to the catalytically important S-D-N loop might explain the kinetic differences between the type A and type D *S. aureus*  $\beta$ -lactamases. Preliminary experiments in which residue 128 has been replaced by other amino acids also have been shown to affect the kinetics of hydrolysis (unpublished observations).

In conclusion, naturally occurring type A, C, and D variants of *S. aureus*  $\beta$ -lactamase exhibit kinetic differences due to sin-

gle amino acid differences at positions close to the active site which have not previously been shown to be involved in enzymatic activity. It is likely that the substitution at the amino acid 128 affects enzyme function by altering the structure of catalytically important S-D-N loop. The differences between the type A and type C enzyme could be due to steric hindrance to substrate binding and/or some structural stabilization effects. This has to be verified by kinetic studies with mutant enzymes in which amino acids 128 and 216 are substituted with different amino acids as well as molecular modelling studies with different  $\beta$ -lactam substrates.

#### ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant AI32126.

We are thankful to K. V. Radha Kishan for providing a ribbon diagram figure (Fig. 1) of *S. aureus*  $\beta$ -lactamase.

#### REFERENCES

- Ambler, R. P. 1980. The structure of beta-lactamases. *Philos. Trans. R. Soc. London B* 289:321-331.
- Ambler, R. P., A. F. W. Coulson, J. M. Frère, J. M. Ghuyen, B. Joris, M. Forsman, R. C. Levesque, G. Tiraby, and S. G. Waley. 1991. A standard numbering scheme for the class A beta-lactamases. *Biochem. J.* 276:269-270.
- Bonfiglio, G., and D. M. Livermore. 1994. Beta-lactamase types amongst *Staphylococcus aureus* isolates in relation to susceptibility to beta-lactamase inhibitor combinations. *J. Antimicrob. Chemother.* 33:465-481.
- Bush, K., G. A. Jacoby, and A. A. Medeiros. 1995. A functional classification scheme for  $\beta$ -lactamases and its correlation with molecular structure. *Antimicrob. Agents Chemother.* 39:1211-1233.
- Chan, P. T. 1986. Nucleotide sequence of the *Staphylococcus aureus* PC1 beta-lactamase gene. *Nucleic Acids Res.* 14:5940.
- Chang, S., and S. N. Cohen. 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol. Gen. Genet.* 168:111-115.
- Delaire, M., R. Labia, J. P. Samama, and J. M. Masson. 1992. Site-directed mutagenesis at the active site of *Escherichia coli* TEM-1 beta-lactamase. Suicide inhibitor-resistant mutants reveal the role of arginine 244 and methionine 69 in catalysis. *J. Biol. Chem.* 267:20600-20606.
- East, A. K., S. P. Curnock, and K. G. H. Dyke. 1990. Change of a single amino acid in the leader peptide of a staphylococcal beta-lactamase prevents the appearance of the enzyme in the medium. *FEMS Microbiol. Lett.* 69: 249-254.
- East, A. K., and K. G. H. Dyke. 1989. Cloning and sequence determination of six *Staphylococcus aureus* beta-lactamases and their expression in *Escherichia coli* and *Staphylococcus aureus*. *J. Gen. Microbiol.* 135:1001-1015.
- Galletto, D. W., J. L. Johnston, and G. L. Archer. 1987. Molecular epidemiology of trimethoprim resistance among coagulase-negative staphylococci. *Antimicrob. Agents Chemother.* 31:1683-1688.
- Herzberg, O. 1991. Refined crystal structure of beta-lactamase from *Staphylococcus aureus* PC1 at 2.0 Å resolution. *J. Mol. Biol.* 217:701-719.
- Herzberg, O., and J. Moul. 1987. Bacterial resistance to beta-lactam antibiotics: crystal structure of beta-lactamase from *Staphylococcus aureus* PC1 at 2.5 Å resolution. *Science* 236:694-701.
- Horinouchi, S., and B. Weisblum. 1982. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. *J. Bacteriol.* 150:804-814.
- Huletsky, A., J. R. Knox, and R. C. Levesque. 1993. Role of Ser-238 and Lys-240 in the hydrolysis of third-generation cephalosporins by SHV-type beta-lactamases probed by site-directed mutagenesis and three-dimensional modeling. *J. Biol. Chem.* 268:3690-3697.
- Imitiaz, U., E. Billings, J. R. Knox, E. K. Manavanth, S. A. Lerner, and S. Mobashery. 1993. Inactivation of class A  $\beta$ -lactamases by clavulanic acid: the role of arginine-244 in a proposed nonconcerted sequence of events. *J. Am. Chem. Soc.* 115:4435-4442.
- Jacob, F., B. Joris, S. Lepage, J. Dusart, and J. M. Frère. 1990. Role of the conserved amino acids of the "SDN" loop (Ser130, Asp131 and Asn132) in a class A beta-lactamase studied by site-directed mutagenesis. *Biochem. J.* 271:399-406.
- Jacoby, G. A., and A. A. Medeiros. 1991. More extended-spectrum  $\beta$ -lactamases. *Antimicrob. Agents Chemother.* 35:1697-1704.
- Johnston, L. H., and K. G. H. Dyke. 1971. Stability of penicillinase plasmids in *Staphylococcus aureus*. *J. Bacteriol.* 107:68-73.
- Kernodle, D. S., P. A. McGraw, C. W. Stratton, and A. B. Kaiser. 1990. Use of extracts versus whole-cell bacterial suspensions in the identification of *Staphylococcus aureus*  $\beta$ -lactamase variants. *Antimicrob. Agents Chemother.* 34:420-425.
- Kernodle, D. S., C. W. Stratton, L. W. McMurray, J. R. Chipley, and P. A.

- McGraw. 1989. Differentiation of beta-lactamase variants of *Staphylococcus aureus* by substrate hydrolysis profiles. *J. Infect. Dis.* **159**:103–108.
21. Kernodle, D. S., D. J. Zygmunt, P. A. McGraw, and J. R. Chipley. 1990. Purification of *Staphylococcus aureus*  $\beta$ -lactamases by using sequential cation-exchange and affinity chromatography. *Antimicrob. Agents Chemother.* **34**:2177–2183.
  22. Kraulis, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **24**:946–950.
  23. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488–492.
  24. Kunkel, T. A., J. D. Roberts, and R. A. Zakour. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**:367–382.
  25. Lacey, R. W., and J. Grinstead. 1972. Linkage of fusidic acid resistance to the penicillinase plasmid in *Staphylococcus aureus*. *J. Gen. Microbiol.* **73**:501–508.
  26. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
  27. Medeiros, A. A. 1984. Beta-lactamases. *Br. Med. Bull.* **40**:18–27.
  28. Miller, S., J. Janin, A. M. Lesk, and C. Chotia. 1987. Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**:641–656.
  29. Novick, R. P. 1963. Analysis by transduction of mutations affecting penicillinase formation in *Staphylococcus aureus*. *J. Gen. Microbiol.* **33**:121–136.
  30. Osuna, J., H. Viadiu, A. L. Fink, and X. Soberón. 1995. Substitution of Asp for Asn at position 132 in the active site of TEM beta-lactamase. Activity toward different substrates and effects of neighboring residues. *J. Biol. Chem.* **270**:775–780.
  31. Paul, G. C., A. B. Gerbaud, A. M. Philippon, B. Pangon, and P. Courvalin. 1989. TEM-4, a new plasmid-mediated  $\beta$ -lactamase that hydrolyzes broad-spectrum cephalosporins in a clinical isolate of *Escherichia coli*. *Antimicrob. Agents Chemother.* **33**:1958–1963.
  32. Petit, A., D. L. Sirot, J. L. Chanal, R. L. Sirot, G. Gerbaud, and R. A. Cluzel. 1989. Molecular epidemiology of TEM-3 (CTX-1)  $\beta$ -lactamase. *Antimicrob. Agents Chemother.* **32**:626–630.
  33. Richardson, J. S., and D. C. Richardson. 1989. In Gerald D. Fasman (ed.), Prediction of protein structure and principles of protein conformation, p. 64. Plenum Press, New York.
  34. Richmond, M. H. 1965. Wild type variants of exopenicillinase from *Staphylococcus aureus*. *Biochem. J.* **88**:452–459.
  35. Rosdahl, V. T. 1973. Naturally occurring constitutive  $\beta$ -lactamase with novel serotype in *Staphylococcus aureus*. *J. Gen. Microbiol.* **77**:229–231.
  36. Rosdahl, V. T. 1986. Penicillinase production in *Staphylococcus aureus* strains of clinical importance. *Dan. Med. Bull.* **33**:175–184.
  37. Ross, G. W., K. V. Chanter, A. M. Harris, S. M. Kirby, M. J. Marshall, and C. H. O'Callaghan. 1974. Comparison of assay techniques for beta-lactamase activity. *Anal. Biochem.* **54**:9–16.
  38. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
  39. Waley, S. G. 1974. A spectrophotometric assay of beta-lactamase action on penicillins. *Biochem. J.* **139**:789–790.
  40. Wang, P., and R. P. Novick. 1987. Nucleotide sequence and expression of the  $\beta$ -lactamase gene from *Staphylococcus aureus* plasmid p1258 in *Escherichia coli*, *Bacillus subtilis*, and *Staphylococcus aureus*. *J. Bacteriol.* **169**:1763–1766.
  41. Wong, J. T. 1985. Kinetics of enzyme mechanisms. Academic Press, Inc., New York.
  42. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.
  43. Zygmunt, D. J., C. W. Stratton, and D. S. Kernodle. 1992. Characterization of four  $\beta$ -lactamases produced by *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **36**:440–445.

## Construction and Characterization of Mutants of the TEM-1 $\beta$ -Lactamase Containing Amino Acid Substitutions Associated with both Extended-Spectrum Resistance and Resistance to $\beta$ -Lactamase Inhibitors

PAUL D. STAPLETON,\* KEVIN P. SHANNON, AND GARY L. FRENCH

Department of Microbiology, The Guy's, King's & St. Thomas' School of Medicine,  
St. Thomas' Campus, London SE1 7EH, United Kingdom

Received 5 October 1998/Returned for modification 25 January 1999/Accepted 12 May 1999

Extended-spectrum TEM  $\beta$ -lactamases (ESBLs) do not usually confer resistance to  $\beta$ -lactamase inhibitors such as clavulanate or tazobactam. To investigate the compatibility of the two phenotypes we used site-directed mutagenesis of the *bla*<sub>TEM-1</sub> gene to introduce into the TEM-1  $\beta$ -lactamase amino acid substitutions that confer the ESBL phenotype: TEM-12 (Arg164→Ser), TEM-26 (Arg164→Ser plus Glu104→Lys), TEM-19 (Gly238→Ser), and TEM-15 (Gly238→Ser plus Glu104→Lys). These were combined with three sets of substitutions that confer inhibitor resistance: TEM-31 (Arg244→Cys), TEM-33 (Met69→Leu), and TEM-35 (Met69→Leu and Asn276→Asp). Introduction of the Arg244→Cys substitution gave rise to inhibitor-resistant hybrid enzymes that either lost ESBL activity (TEM-12, TEM-15, and TEM-19) or had reduced activity (TEM-26) against ceftazidime. In contrast, the introduction of Met69→Leu or Met69→Leu plus Asn276→Asp substitutions did not significantly affect the abilities of the enzymes to confer resistance to ceftazidime, although increased susceptibility to cefotaxime was observed with *Escherichia coli* strains that expressed the TEM-19 and TEM-26  $\beta$ -lactamases. With the exception of the TEM-12  $\beta$ -lactamase, introduction of the Met69→Leu substitution did not give rise to enzymes with increased resistance to clavulanate compared to that of the TEM-1  $\beta$ -lactamase. However, introduction of the double substitution Met69→Leu plus Asn276→Asp in the ESBLs did give rise to low-level (TEM-19, TEM-15, and TEM-26) or moderate-level (TEM-12) clavulanate resistance. None of the hybrid enzymes were as resistant to clavulanate as the corresponding inhibitor-resistant TEM  $\beta$ -lactamase mutant, suggesting that active-site configuration in the ESBLs limits the degree of clavulanate resistance conferred.

Gram-negative bacteria may exhibit reduced susceptibility to  $\beta$ -lactam antibiotics by a number of mechanisms including reduced outer membrane permeability, target-site modification, and efflux of the  $\beta$ -lactam out of the cell (20, 23). However, by far the most common mechanism of resistance is the enzymatic inactivation of the  $\beta$ -lactam by a  $\beta$ -lactamase (18). There are many types of  $\beta$ -lactamases, which have been classified by their amino acid sequences and corresponding substrate profiles (6). The TEM-1  $\beta$ -lactamase belongs to a functional group of broad-spectrum enzymes that are inhibited by clavulanate (6). This group includes enzymes such as the SHV-1 and OHIO-1  $\beta$ -lactamases. Although the TEM-1  $\beta$ -lactamase does not usually provide protection against extended-spectrum cephalosporins such as ceftazidime and cefotaxime or  $\beta$ -lactamase inhibitors like clavulanate and tazobactam (except in the case of TEM-1 overproduction), amino acid substitutions can alter the hydrolytic spectrum of the  $\beta$ -lactamase to encompass these compounds.

Extended-spectrum TEM  $\beta$ -lactamases (ESBLs) do not usually confer resistance to  $\beta$ -lactamase inhibitors, suggesting that the two phenotypes may be incompatible. In support of this suggestion, Imtiaz et al. (15) have shown that introduction of an amino acid substitution (Arg164→Ser) that confers on the TEM-1  $\beta$ -lactamase the ability to efficiently hydrolyze ceftazi-

dime leads to the loss of clavulanate resistance when introduced into the inhibitor-resistant  $\beta$ -lactamase TEM-31. However, recently a clinical *Escherichia coli* isolate that expressed a  $\beta$ -lactamase, TEM-50 (CMT-1), that conferred low-level resistance both to  $\beta$ -lactamase inhibitors and to extended-spectrum cephalosporins has been reported (22).

In order to investigate this phenomenon further we used site-directed mutagenesis of the TEM  $\beta$ -lactamase encoding gene to introduce into ESBLs amino acid substitutions known to confer inhibitor resistance. We found that the different amino acid substitutions gave rise to enzymes that conferred different resistance phenotypes. None of the substitutions conferred high-level resistance to both  $\beta$ -lactamase inhibitors and extended-spectrum cephalosporins, although the double amino acid substitution (Met69→Leu, Asn276→Asp) in the TEM-12  $\beta$ -lactamase did give rise to an ESBL with a moderate level of clavulanate resistance.

### MATERIALS AND METHODS

**Bacterial strains and plasmids.** *E. coli* CJ236 [*dut-1 ung-1 thi-1 relA1*; pCJ105 (Cm<sup>r</sup>)] and *E. coli* MV1190 [ $\Delta$ (*lac-proAB*) *thi supE*  $\Delta$ (*srl-recA*)306::Tn10(Tet<sup>r</sup>); (F'*traD36 proAB lac*<sup>r</sup>Z $\Delta$ M15)] were used in this study. The plasmid vector pTZ18U was used as the initial source of the *bla*<sub>TEM</sub> gene. All bacteria were grown in Luria-Bertani (LB) broth or on LB agar (Oxoid, Basingstoke, United Kingdom) containing the appropriate antibiotic (chloramphenicol, 20  $\mu$ g/ml; amoxicillin, 100  $\mu$ g/ml; or tetracycline, 10  $\mu$ g/ml).

**Antibiotics and reagents.** The following companies kindly supplied antibiotic powders of known potencies: Bristol Meyers Squibb (cefepime and aztreonam); American Cyanamid (piperacillin and tetracycline); Glaxo Group Research Ltd. (ceftazidime and cephaloridine); Roussel Laboratories Ltd. (cefotaxime and chloramphenicol); and SmithKline Beecham (amoxicillin, clavulanate, temocillin, and ticarcillin). Nitrocefin was obtained from Oxoid.

\* Corresponding author. Mailing address: Department of Biology (Darwin Building), University College London, Gower Street, London WC1E 6BT, United Kingdom. Phone: 44 171 504 2934. Fax: 44 171 380 7098. E-mail: p.stapleton@ucl.ac.uk.



TABLE 1. Oligonucleotides used in site-directed mutagenesis experiments and in DNA sequencing

Procedure and oligonucleotide	Sequence (5'-3')	Codon change and nucleotide position <sup>b</sup>
<b>Mutagenesis</b>		
Ile86→Val	GGCGTCAACACGGGATAAT	ATT→GTT
Val184→Ala	ATTGCGGCAGGCATCGTGG	GTA→GCC
Met69→Leu	TAAAGTGCTCAGCATTGGAAC	ATG→CTG
Asn276→Asp	TCTGTCTATCTCGTTCATCC	AAT→GAT
Arg244→Cys	ATGATACCGCAAGACCC	CGC→TGC
Glu104→Lys	GTGAGTATTTAACCAAGTC	GAG→AAA
Gly238→Ser	CACGCTCACIGGCTCCAGATTTAT	GGT→AGT
Arg164→Ser	GTTCCCAAGAATCAAGGC	CGT→TCT
<b>Sequencing</b>		
f-TEM2F	GTATGAGTATTCAACATTTCCGTGTCG	205-231
f-TEM2R	ACCAATGCTTAATCAGTGAGGCA	1064-1042
f-TEMi	ACTGTCATGCCATCCGTAAGA	556-536
f-TEM2i	CTGCGGCCAACTTACTTCTGACAA	598-621

<sup>a</sup> Altered bases are underlined.<sup>b</sup> Nucleotide positions are according to Sutcliffe (27).

**Susceptibility testing.** MICs were determined by agar dilution on Diagnostic Sensitivity Test Agar (CM261; Oxoid) with an inoculum of about 10<sup>4</sup> organisms per spot as described previously (24). *E. coli* NCTC 10418 was used as the control strain.

**Site-directed mutagenesis.** Site-directed mutagenesis was performed with the reagents contained within the Muta-Gene Phagemid In Vitro Mutagenesis kit (version 2) from Bio-Rad (Hemel Hempstead, United Kingdom). The procedures used in this kit are based on the method originally described by Kunkel et al. (17). Oligonucleotides were designed with the aid of oligonucleotide design software (PrimerSelect; DNAStar) and were based on the sequence of the *bla*<sub>TEM-1</sub> gene reported by Sutcliffe (27). The oligonucleotides were custom made by Pharmacia Biotech (St. Albans, United Kingdom) (Table 1).

**DNA sequencing.** In order to confirm that mutations had been introduced, plasmid DNA was extracted with a Qiagen QIAprep kit (Qiagen Ltd., Crawley, United Kingdom) and was sequenced in both directions with fluorescein-labelled primers (Table 1). DNA sequencing was performed with the reagents contained in a cycle sequencing kit (RPN 2438) from Amersham Life Sciences (Little Chalfont, United Kingdom) by following the manufacturer's instructions. The annealing temperature for the cycle sequencing reactions was 60°C, and the DNA sequence was determined with an automated DNA sequencer (Pharmacia Biotech).

**Determination of IC<sub>50</sub>s.** Each strain was grown at 37°C in brain heart infusion broth (Oxoid) for 16 h, with shaking (200 rpm). The cells were harvested by centrifugation and were resuspended in 0.5 ml of sterile distilled water, and the β-lactamase was released by sonication. Sonication was performed for 20 s with a W-385 sonicator (Heat Systems; Ultrasonics, Inc., Farmingdale, N.Y.) with the following settings: 5-s cycle time, 50% duty cycle, and a 1.5 output control setting. β-Lactamase activity was measured by monitoring the rate of nitrocefin hydrolysis (10 μM) at 482 nm in a Biochrom 4060 spectrophotometer (Pharmacia Biotech). All assays were performed in 0.1 M phosphate buffer (pH 7.0) and at 37°C. In order to take into account the different levels of β-lactamase activities within the samples the activity of each sample was standardized to give an absorbance change of 0.15 per min. Samples were preincubated for 10 min at 37°C with various concentrations (0.01 to 50 μM) of the β-lactamase inhibitor before the β-lactamase activity was determined with nitrocefin (10 μM) as the reporter substrate. The concentration of β-lactamase inhibitor required to inhibit 50% of the β-lactamase activity (IC<sub>50</sub>) was then determined graphically.

## RESULTS

**Mutagenesis.** Phagemid pTZ18U conveniently encodes a *bla*<sub>TEM</sub> gene, and use of pTZ18U thus negates the need to subclone the *bla*<sub>TEM</sub> gene from another source. However, the *bla*<sub>TEM</sub> gene from pTZ18U is not identical to *bla*<sub>TEM-1</sub> as the result of two nucleotide changes, G<sup>244</sup>→A and C<sup>545</sup>→T, that were introduced to remove *Pst*I and *Hinc*II restriction sites, respectively. While the resulting amino acid substitutions, Ile84→Val and Ala184→Val, have been regarded as neutral (22a), Chaibi et al. (8) have demonstrated that the catalytic efficiency of the "artificial" TEM β-lactamase was one-half to one-third lower than that of the TEM-1 β-lactamase. Consequently, in this study we initially converted the artificial *bla*<sub>TEM</sub>

into *bla*<sub>TEM-1</sub> and subsequently used this gene as the template for the construction of the TEM mutants. Four ESBL enzymes (TEM-12, TEM-15, TEM-19, and TEM-26) were constructed together with three β-lactamase-inhibitor-resistant mutants (TEM-31, TEM-33, and TEM-35) (Table 2). In order to investigate whether the amino acid substitutions found in β-lactamase inhibitor-resistant mutants could confer inhibitor resistance if introduced into ESBL enzymes, the three sets of amino acid substitutions that confer inhibitor resistance were engineered into the extended-spectrum antibiotic-resistant TEM β-lactamases by altering the gene-coding sequence. The amino acid substitutions corresponded to those found in the TEM-31 (Arg244→Cys), TEM-33 (Met69→Leu), and TEM-35 (Met69→Leu and Asn276→Asp) β-lactamases. In all cases the introduced nucleotide changes in the *bla*<sub>TEM</sub> gene were confirmed by DNA sequencing.

**Phenotypic characterization of TEM-1 β-lactamase and mutant derivatives.** (i) **TEM-1 and ESBL enzymes.** The MICs of ampicillin and ticarcillin in the presence of clavulanate (2 μg/ml) and piperacillin in the presence of tazobactam (4 μg/ml) for *E. coli* MV1190 expressing the TEM-1 β-lactamase were relatively high (Table 2). This could be accounted for by the large quantity of the TEM-1 β-lactamase expressed as a result of the high copy number of the pTZ18U plasmid carrying the *bla*<sub>TEM-1</sub> gene (Table 3). Despite this, because the TEM-1 β-lactamase and the mutant enzymes in this study shared the same genetic background, comparisons between the mutant enzymes and the TEM-1 β-lactamase could still be made.

The TEM-12, TEM-15, and TEM-26 β-lactamases were found to confer 16- to 128-fold higher levels of resistance to ceftazidime than the TEM-1 β-lactamase, confirming that these enzymes were indeed ESBLs (Table 2). Although the TEM-19 β-lactamase did not confer increased levels of resistance to ceftazidime, a 16-fold increase in the level of resistance to cefotaxime was observed. Cefepime was found to be less effective against *E. coli* MV1190 strains that expressed the TEM-12 and TEM-26 β-lactamases, and with the exception of TEM-19, the ESBLs conferred higher levels of resistance to aztreonam than the TEM-1 β-lactamase did. *E. coli* MV1190 expressing either of the four ESBL enzymes was found to be more susceptible to the penicillin-β-lactamase inhibitor combinations than *E. coli* MV1190 expressing the TEM-1 β-lactamase. None of the ESBL enzymes conferred increased resistance to temocillin. Measurement of the β-lactamase activities

TABLE 2. MICs of  $\beta$ -lactams and  $\beta$ -lactamase inhibitor combinations for *E. coli* MV1190 producing the TEM-1 and mutant TEM  $\beta$ -lactamases

Strain or amino acid changes from TEM-1 <sup>a</sup>	Designation	MIC (μg/ml) <sup>b</sup>													
		AMX	AMX + CLA <sup>c</sup>	TIC	TIC + CLA <sup>c</sup>	PIP	PIP + TZB <sup>d</sup>	CLD	CTX	CAZ	CAZ + CLA <sup>c</sup>	FEP	AZM	TEM	
TEM-1 <sup>e</sup>	TEM-1	>1,024	256	>1,024	256	128	32	64	0.06	0.5	0.25	0.12	0.25	4	
M69L	TEM-33	>1,024	>1,024	>1,024	1,024	128	64	16	0.06	0.5	0.25	≤0.06	0.25	4	
M69L, N276D	TEM-35	>1,024	>1,024	1,024	512	128	32	16	0.06	0.25	0.25	≤0.06	0.12	4	
R244C	TEM-31	1,024	512	64	64	8	4	2	0.06	0.25	0.12	≤0.06	0.12	4	
G238S	TEM-19	>1,024	16	1,024	16	32	≤1	32	1	0.5	0.25	0.12	0.5	4	
G238S, M69L		>1,024	32	512	16	32	≤1	32	0.25	0.25	0.25	0.12	0.25	8	
G238S, M69L, N276D		>1,024	128	512	32	32	≤1	32	0.12	0.25	0.25	0.25	0.12	4	
G238S, R244C		1,024	128	64	16	8	2	4	0.06	0.25	0.25	≤0.06	0.12	4	
G238S, E104K	TEM-15	>1,024	16	1,024	32	32	≤1	32	8	8	0.25	0.25	4	8	
G238S, E104K, M69L		>1,024	32	1,024	64	32	≤1	32	4	4	0.25	0.25	1	4	
G238S, E104K, M69L, N276D	TEM-50	>1,024	128	1,024	128	32	≤1	32	4	4	0.5	0.5	2	4	
G238S, E104K, R244C		1,024	64	256	64	8	≤1	4	0.06	0.25	0.25	0.25	1	4	
R164S	TEM-12	>1,024	32	1,024	32	64	2	32	0.12	8	0.25	1	1	8	
R164S, M69L		>1,024	256	512	64	32	≤1	8	0.12	8	0.25	0.5	1	4	
R164S, M69L, N276D		>1,024	1,024	512	128	32	2	8	0.06	8	2	0.5	0.5	4	
R164S, R244C		1,024	256	16	8	4	2	2	0.06	0.25	0.25	≤0.06	0.12	4	
R164S, E104K	TEM-26	>1,024	16	>1,024	32	64	≤1	8	1	64	0.5	1	16	8	
R164S, E104K, M69L		>1,024	64	1,024	64	64	2	8	0.5	64	2	1	16	4	
R164S, E104K, M69L, N276D		>1,024	256	512	128	32	2	8	0.12	64	8	1	4	4	
R164S, E104K, R244C		512	64	128	32	32	8	2	0.06	4	1	0.25	1	8	
E. coli MV1190 (recipient strain)		8	4	2	2	≤1	≤1	2	0.06	0.12	0.12	≤0.06	0.12	4	

<sup>a</sup> M69L refers to an amino acid substitution of leucine for methionine at position 69 in the TEM-1  $\beta$ -lactamase protein sequence. The single amino acid codes for the other substitutions are as follows: N, asparagine; D, aspartate; G, glycine; R, arginine; C, cysteine; S, serine; E, glutamate; and K, lysine.

<sup>b</sup> AMX, amoxicillin; CLA, clavulanate; TIC, ticarcillin; PIP, piperacillin; TZB, tazobactam; CLD, cephaloridine; CTX, cefotaxime; CAZ, ceftazidime; FEP, cefepime; AZM, aztreonam; TEM, temocillin.

<sup>c</sup> Fixed concentration of clavulanate (2  $\mu$ g/ml).

<sup>d</sup> Fixed concentration of tazobactam (4  $\mu$ g/ml).

<sup>e</sup> TEM with amino acid modifications engineered to be identical to natural TEM-1.

TABLE 3.  $\beta$ -Lactamase activities and clavulanate and tazobactam  $IC_{50}$ s for TEM-1  $\beta$ -lactamase and TEM mutant enzymes

Amino acid change from TEM-1 <sup>a</sup>	Designation	$\beta$ -Lactamase activity <sup>b</sup>	$IC_{50}$ ( $\mu$ M)	
			Clavulanate	Tazobactam
TEM-1	TEM-1	5,330	0.08	0.03
M69L	TEM-33	586	2	0.5
M69L, N276D	TEM-35	1,190	12	0.6
R244C	TEM-31	514	10	1.7
G238S	TEM-19	2,950	0.001	0.002
G238S, M69L		853	0.03	0.02
G238S, M69L, N276D		2,360	0.38	0.04
G238S, R244C		485	1.5	0.3
G238S, E104K	TEM-15	29	0.002	0.004
G238S, E104K, M69L		319	0.07	0.01
G238S, E104K, M69L, N276D	TEM-50	952	0.3	0.02
G238S, E104K, R244C		696	1.1	1.2
R164S	TEM-12	25	0.02	0.02
R164S, M69L		49	0.25	0.2
R164S, M69L, N276D		1,440	1.8	0.3
R164S, R244C		937	4.5	5
R164S, E104K	TEM-26	523	<0.01	0.06
R164S, E104K, M69L		211	0.08	0.3
R164S, E104K, M69L, N276D		458	0.35	0.2
R164S, E104K, R244C		54	3.5	2

<sup>a</sup> See footnote a to Table 2 for a key to the amino acid substitutions.<sup>b</sup> Activities are expressed as nanomoles of nitrocefin hydrolyzed per minute per milligram of protein.

of the ESBLs with nitrocefin as the reporter substrate indicated that the ESBL enzymes had lower levels of activity against nitrocefin than the TEM-1  $\beta$ -lactamase; this was also true for the other mutant TEM  $\beta$ -lactamases (Table 3).

(ii) **Inhibitor-resistant enzymes.** The substitutions Arg244→Cys, Met69→Leu, and Met69→Leu plus Asn276→Asp in the TEM-1  $\beta$ -lactamase gave rise to enzymes with resistance to clavulanate combined with resistance to amoxicillin or ticarcillin. Substitution of a Cys residue at position 244 of the TEM-1  $\beta$ -lactamase also resulted in an enzyme (TEM-31) that conferred lower levels of resistance to penicillins and cephaloridine than the levels conferred by TEM-1 (Table 2). In contrast, a single Met69→Leu substitution and a double substitution, Met69→Leu plus Asn276→Asp, in the TEM-1  $\beta$ -lactamase did not greatly affect the MICs of piperacillin, although a fourfold reduction in resistance to cephaloridine was observed (Table 2). None of the inhibitor-resistant enzymes conferred resistance to extended-spectrum cephalosporins, aztreonam, or temocillin.

(iii) **Substitution of Cys for Arg at position 244 in ESBLs.** Introduction of the Arg244→Cys substitution into the ESBL enzymes had an effect similar to that in TEM-1. Like the TEM-31  $\beta$ -lactamase, the resulting hybrid enzymes conferred lower levels of resistance to penicillins and cephaloridine than their respective parent enzymes did. However, the MICs of

penicillin-inhibitor combinations were elevated for the strain with TEM-26 plus the Arg244→Cys substitution. In addition, the amino acid substitution resulted in enzymes that either had lost (TEM-12, TEM-15, and TEM-19) or had a reduced ability (TEM-26) to confer resistance to ceftazidime. The MICs of cefepime and aztreonam were reduced 4- and 16-fold, respectively, for *E. coli* MV1190 expressing the hybrid TEM-26  $\beta$ -lactamase compared to the MICs for the strain expressing the TEM-1  $\beta$ -lactamase. All the hybrid enzymes with the Arg244→Cys substitution were found to be more resistant to clavulanate and tazobactam inhibition than the TEM-1  $\beta$ -lactamase, with the  $IC_{50}$ s for the hybrid enzymes being comparable to those for the naturally occurring inhibitor-resistant TEM  $\beta$ -lactamases (Table 3).

(iv) **Substitution of Leu for Met at position 69 in ESBLs.** The  $IC_{50}$ s of clavulanate for the parental ESBL enzymes were lower than those of the TEM-1  $\beta$ -lactamase, indicating that the ESBL enzymes were more susceptible to clavulanate inhibition than the TEM-1  $\beta$ -lactamase. Introduction of a Leu residue at position 69 in the ESBLs resulted in hybrid enzymes that conferred increased levels of resistance to both clavulanate and tazobactam compared to the level of resistance conferred by their respective parental ESBL enzymes (Table 3). In the case of the TEM-12  $\beta$ -lactamase, the amino acid substitution gave rise to a hybrid enzyme that was less susceptible to clavulanate inhibition than the TEM-1  $\beta$ -lactamase. For the other ESBLs, however, the substitution resulted in hybrid enzymes for which clavulanate  $IC_{50}$ s were similar to those for the TEM-1  $\beta$ -lactamase. Tazobactam was found to be equally effective against the TEM-1  $\beta$ -lactamase and the hybrid ESBLs with a Gly238→Ser substitution (TEM-15 and TEM-19). However, tazobactam was less effective against hybrid ESBLs with the Arg164→Ser substitution (TEM-12 and TEM-26) (Table 3).

In contrast to the Arg244→Cys substitution, introduction of a Leu residue at position 69 in the ESBLs resulted in hybrid enzymes that retained the ability to confer resistance to ceftazidime and, in the case of the TEM-12 hybrid enzyme, that had increased levels of resistance to ceftazidime in combination with clavulanate. However, the amino acid substitution in the TEM-19  $\beta$ -lactamase gave rise to a hybrid enzyme that conferred a lower level of resistance to cefotaxime than the parent enzyme did.

(v) **Substitution of Leu for Met at position 69 and Asp for Asn at position 276 in ESBLs.** Introduction of the double amino acid substitution Leu-69 and Asp-276 in the ESBLs gave rise to hybrid enzymes that were more resistant to clavulanate inhibition than hybrid ESBL enzymes with a single Leu-69 substitution. In the case of the TEM-12  $\beta$ -lactamase, this gave rise to an enzyme for which the  $IC_{50}$  of clavulanate was similar to that for the inhibitor-resistant enzyme TEM-33. For the other hybrid ESBLs, however, the  $IC_{50}$ s of clavulanate were intermediate between that for the TEM-1  $\beta$ -lactamase and those for the inhibitor-resistant enzymes. The  $IC_{50}$ s of tazobactam were similar for the hybrid enzymes with single or double amino acid substitutions. The double substitution did not greatly affect the ability of the enzymes to confer resistance to ceftazidime, although increased susceptibility to cefotaxime was apparent with *E. coli* MV1190 expressing the hybrid derivatives of the TEM-19 and TEM-26  $\beta$ -lactamases. Two of the hybrid enzymes, TEM-12 and TEM-26, showed a markedly reduced susceptibility to ceftazidime combined with their susceptibility to clavulanate. As a consequence of the double amino acid substitution, extended-spectrum resistant variants of TEM-12, TEM-15, and TEM-26  $\beta$ -lactamases that also conferred increased levels of resistance to  $\beta$ -lactamase inhibitors were constructed. However, the levels of clavulanate resistance

conferred by the hybrid ESBLs were not as high as that conferred by the corresponding inhibitor-resistant TEM  $\beta$ -lactamase TEM-35.

## DISCUSSION

**Substitution of Cys for Arg at position 244.** In this study we replaced the Arg at position 244 in the TEM-12, TEM-15, TEM-19, and TEM-26  $\beta$ -lactamases with a Cys residue in order to investigate whether the amino acid substitution would give rise to inhibitor-resistant ESBLs. In each case the substitution conferred increased levels of resistance to  $\beta$ -lactamase inhibitors, but the substitution also gave rise to enzymes that conferred lower degrees of resistance to penicillin and cephalosporins. Both these observations are consistent with a disruption of the Arg244 hydrogen-bonding arrangement predicted to occur in the TEM-1  $\beta$ -lactamase (26, 30). Since the Cys residue at position 244 would be unable to form a hydrogen bond to the common carboxylate group of  $\beta$ -lactam antibiotics, this probably explains why the MICs of both penicillins and cephalosporins were affected by the amino acid substitution. This would be especially pertinent if, as suggested by Zafaralla et al. (30), the binding energy of Arg244 is used to lower the activation energy of the hydrolytic reaction.

The resistance to  $\beta$ -lactamase inhibitors conferred by the hybrid enzymes in this study is understandable in light of the essential role that the Arg244 residue plays in maintaining in position the water molecule (Wat399) believed to be important in the inactivation of  $\beta$ -lactamase by clavulanate (14, 28). In naturally occurring variants of the TEM-1 and TEM-2  $\beta$ -lactamases, as in our mutants, substitution of Cys, Ser, or His residues at position 244 has given rise to inhibitor-resistant enzymes (1, 3, 4, 29). The shorter side chains of the substituted amino acids in the inhibitor-resistant variants are thought to be unable to form a hydrogen bond with Wat399, which is displaced as a consequence and which is unable to act as a proton source in the inactivation process (14, 16, 19). However, our results contrast with those of Imtiaz et al. (15), who reported that a substitution of a Ser for Arg at position 244 in the TEM-12  $\beta$ -lactamase (also derived from TEM-1) neither conferred inhibitor resistance nor significantly affected the enzyme's ability to hydrolyze ceftazidime. Why the two different amino acid substitutions gave rise to two different effects is not clear. Imtiaz et al. (15) have suggested that an alteration of the topology of the active site that is caused by the Arg164→Ser substitution in the TEM-12  $\beta$ -lactamase may have resulted in a different clavulanate binding arrangement that promoted a repositioning of the water molecule close to the site of inactivation. Consistent with this suggestion we found that the four ESBLs in this study were more sensitive to clavulanate inhibition than the TEM-1  $\beta$ -lactamase. If a different clavulanate binding arrangement does occur in the hybrid enzymes, the results of this study show that the nature of the residue at position 244 is still important in dictating whether the enzyme is resistant to clavulanate or not. Thus, it would appear that the Ser residue, but not the Cys residue, either performs a role similar to that of the Arg residue in the hybrid enzymes or, through structural rearrangement, promotes another residue to perform a similar function.

**Substitution of Leu for Met at position 69.** Unlike the TEM-1  $\beta$ -lactamase, in which substitutions of Leu, Val, or Ile for the Met at position 69 have all given rise to inhibitor-resistant enzymes (9, 11, 25, 31), substitution of a Leu residue for the Met residue at this position in the four ESBLs did not give rise to clavulanate-resistant enzymes. This probably can be explained by a different binding arrangement of the clavu-

lanate molecule in the active site of the hybrid ESBL enzymes compared to that in the TEM-1  $\beta$ -lactamase. As noted previously the parental ESBLs were more sensitive to clavulanate inhibition than the TEM-1  $\beta$ -lactamase, suggesting that alterations within the active site enhanced the inhibitory action of clavulanate. Although the hybrid ESBLs were not resistant to clavulanate, they were less sensitive to clavulanate inhibition than their respective parent enzymes were. The substitutions at position 69 are thought to cause slight alterations to the active-site structure of the TEM-1  $\beta$ -lactamase, resulting in deformation of the oxyanion hole and a less favorable binding orientation of the clavulanate molecule (10). This suggests that the clavulanate molecule still interacts with the oxyanion hole but possibly in a different manner.

Substitutions of Met69→Ile or Met69→Val in the SHV-5  $\beta$ -lactamase and Met69→Ile in an OHIO-1  $\beta$ -lactamase mutant bearing a Gly238→Ser substitution have all given rise to enzymes that were less susceptible to inhibition by clavulanate than their respective parent enzymes (2, 12). These mutant enzymes exhibited reduced penicillinase activity and, in the case of the SHV enzymes, a reduced ability to hydrolyze cephalothin and cefotaxime (2, 12). In contrast, substitution of a Leu residue at position 69 in the extended-spectrum TEM  $\beta$ -lactamases in this study did not significantly affect the ability of the enzymes to confer resistance to penicillins and ceftazidime, although reduced levels of resistance to cefotaxime were noted with the TEM-19 hybrid enzyme. These variations may or may not be related to the different nature of the substituted residues in each case. While all three residues may exert a hydrophobic effect, only the branched residues Val and Ile are thought to produce additional steric constraints. The smaller impact of the Leu69 substitution on the structure of the TEM-32  $\beta$ -lactamase has been used to explain the lower clavulanate  $K_i$  value for this enzyme compared with that for the TEM-1  $\beta$ -lactamase with Met69→Ile or Met69→Val substitutions (9). Whether the substitution of Val or Ile into position 69 of the ESBLs examined in this study would give rise to hybrid enzymes with greater degrees of clavulanate resistance has yet to be determined. However, in light of the reduced penicillinase activity of the SHV and OHIO enzymes, a similar reduction in resistance to penicillins and possibly cephalosporins may also be observed.

**Substitution of Asp for Asn at position 276 plus Leu for Met at position 69.** Amino acid substitutions at position 276 have been found naturally only in combination with changes at position 69 in the TEM  $\beta$ -lactamase (5, 13, 22, 31), although the change can confer inhibitor resistance in the absence of a substitution at position 69 (7, 21, 28). Recently, Sirot et al. (22) have reported on a natural variant of the TEM-15  $\beta$ -lactamase, designated TEM-50, with amino acid substitutions, Met69→Leu and Asn276→Asp, found in the inhibitor-resistant  $\beta$ -lactamase TEM-35. An *E. coli* strain expressing the TEM-50  $\beta$ -lactamase displayed susceptibilities to  $\beta$ -lactams, including ceftazidime and cefotaxime, that were between those for strains expressing the TEM-15 or TEM-35  $\beta$ -lactamases. In our study we artificially constructed the TEM-50  $\beta$ -lactamase together with mutants of the TEM-12, TEM-19, and TEM-26  $\beta$ -lactamases. In contrast to Sirot et al. (22), we found that the MIC of ceftazidime for *E. coli* MV1190 expressing the TEM-50  $\beta$ -lactamase was only twofold lower than the MIC of ceftazidime for the same strain expressing the TEM-15  $\beta$ -lactamase. A possible explanation for this difference may have been the exceptionally high level of  $\beta$ -lactamase expressed from the high-copy-number plasmid pTZ18U harboring the TEM-coding gene used in this study. Such high levels of  $\beta$ -lactamase expression may have

masked small differences in hydrolytic activities between the enzymes.

In agreement with Sirot et al. (22), we found that the TEM-50  $\beta$ -lactamase conferred low levels of resistance to clavulanate. We also demonstrated that when the double amino acid substitutions were introduced into the TEM-19, TEM-12, and TEM-26  $\beta$ -lactamases the resulting enzymes also conferred increased levels of resistance to clavulanate and, in the case of the TEM-12 and TEM-26 derivatives, retained ceftazidime resistance. Indeed, these two hybrid mutants showed considerably reduced levels of susceptibility to the ceftazidime-clavulanate combination. Previous studies have shown that clavulanate is more potent against strains that produce inhibitor-resistant TEM  $\beta$ -lactamases with a single substitution (Asn276 $\rightarrow$ Asp) than against those that have double substitutions (Met69 $\rightarrow$ Leu and Asn276 $\rightarrow$ Asp) (5, 7). Similarly, we demonstrate that double substitutions within the four ESBLs in this study also resulted in hybrid enzymes that conferred greater resistance to clavulanate than the levels of resistance conferred by those with the single Met69 $\rightarrow$ Leu substitution. Furthermore, consistent with the study of Caniça et al. (7) on inhibitor-resistant TEM  $\beta$ -lactamases, we found tazobactam to be more potent than clavulanate against strains producing inhibitor-resistant enzymes with double substitutions. Thus, there appears to be a correlation between the inhibitor resistance phenotypes conferred by the single (Met69 $\rightarrow$ Leu) and double (Met69 $\rightarrow$ Leu and Asn276 $\rightarrow$ Asp) substitutions in the TEM-1  $\beta$ -lactamase and those conferred by the same substitutions in the extended-spectrum TEM  $\beta$ -lactamases.

In conclusion, of the hybrid enzymes constructed, the hybrid of the TEM-12  $\beta$ -lactamase conferred the greatest reduction in sensitivity to clavulanate while it retained the ability to confer resistance to ceftazidime. As with all the hybrid enzymes, including those with the Arg244 $\rightarrow$ Cys substitutions, the level of resistance to penicillin-clavulanate combinations that was conferred (Table 2) and the reduction in the degree of sensitivity to inhibition by clavulanate (Table 3) were not as high as those for equivalent inhibitor-resistant TEM  $\beta$ -lactamases. This suggests that the altered active-site configuration in the ESBL enzymes limits the degree of clavulanate resistance conferred by the ESBL-inhibitor hybrid enzymes. Whether this is due to a different binding arrangement of the clavulanate molecule in the active site of the extended-spectrum TEM  $\beta$ -lactamases or some other factor has yet to be determined.

#### ACKNOWLEDGMENT

This work was funded by a project grant (grant 804) from the Special Trustees of St. Thomas' Hospital.

#### REFERENCES

- Belaouaj, A., C. Lapoumeroulie, M. M. Caniça, G. Vedel, P. Nénot, R. Krishnamoorthy, and G. Paul. 1994. Nucleotide sequences of the genes coding for the TEM-like  $\beta$ -lactamases IRT-1 and IRT-2 (formerly called TRI-1 and TRI-2). *FEMS Microbiol. Lett.* 120:75-80.
- Bonomo, R. A., J. R. Knox, S. D. Rudin, and D. M. Shtaes. 1997. Construction and characterization of an OHIO-1  $\beta$ -lactamase bearing Met69Ile and Gly238Ser mutations. *Antimicrob. Agents Chemother.* 41:1940-1943.
- Bret, L., E. B. Chaibi, C. Chanal-Claris, D. Sirot, R. Labia, and J. Sirot. 1997. Inhibitor-resistant TEM (IRT)  $\beta$ -lactamases with different substitutions at position 244. *Antimicrob. Agents Chemother.* 41:2547-2549.
- Bret, L., C. Chanal, D. Sirot, R. Labia, and J. Sirot. 1996. Characterization of an inhibitor-resistant enzyme IRT-2 derived from TEM-2  $\beta$ -lactamase produced by *Proteus mirabilis* strains. *J. Antimicrob. Chemother.* 38:183-191.
- Brun, T., J. Péduzzi, M. M. Caniça, G. Paul, P. Nénot, M. Barthélémy, and R. Labia. 1994. Characterization and amino acid sequence of IRT-4, a novel TEM-type enzyme with a decreased susceptibility to  $\beta$ -lactamase inhibitors. *FEMS Microbiol. Lett.* 120:111-118.
- Bush, K., G. A. Jacoby, and A. A. Medeiros. 1995. A functional classification scheme for  $\beta$ -lactamases and its correlation with molecular structure. *Antimicrob. Agents Chemother.* 39:1211-1233.
- Caniça, M. M., N. Caroff, M. Barthélémy, R. Labia, R. Krishnamoorthy, G. Paul, and J.-M. Dupret. 1998. Phenotypic study of resistance of  $\beta$ -lactamase-inhibitor-resistant TEM enzymes which differ by naturally occurring variations and by site-directed substitution at Asp<sup>276</sup>. *Antimicrob. Agents Chemother.* 42:1323-1328.
- Chaibi, E. B., J. Péduzzi, M. Barthélémy, and R. Labia. 1997. Are TEM  $\beta$ -lactamases encoded by pBR322 and Bluescript plasmids enzymatically indistinguishable? *J. Antimicrob. Chemother.* 39:668-669.
- Chaibi, E. B., J. Péduzzi, S. Farzaneh, M. Barthélémy, D. Sirot, and R. Labia. 1998. Clinical inhibitor-resistant mutants of the  $\beta$ -lactamase TEM-1 at amino-acid position 69. Kinetic analysis and molecular modelling. *Biochim. Biophys. Acta* 1382:38-46.
- Delaire, M., R. Labia, J. P. Samama, and J. M. Masson. 1992. Site-directed mutagenesis at the active site of *Escherichia coli* TEM-1  $\beta$ -lactamase. Suicide inhibitor-resistant mutants reveal the role of arginine 244 and methionine 69 in catalysis. *J. Biol. Chem.* 267:20600-20606.
- Farzaneh, S., E. B. Chaibi, J. Péduzzi, M. Barthélémy, R. Labia, J. Blazquez, and F. Baquero. 1996. Implication of Ile-69 and Thr-182 residues in kinetic characteristics of IRT-3 (TEM-32)  $\beta$ -lactamase. *Antimicrob. Agents Chemother.* 40:2434-2436.
- Giakkoup, P., V. Miriagou, M. Gazouli, E. Tzelepi, N. J. Legakis, and L. S. Tzouveleki. 1998. Properties of mutant SHV-5  $\beta$ -lactamases constructed by substitution of isoleucine or valine for methionine at position 69. *Antimicrob. Agents Chemother.* 42:1281-1283.
- Henquell, C., C. Chanal, D. Sirot, R. Labia, and J. Sirot. 1995. Molecular characterization of nine different types of mutants among 107 inhibitor-resistant TEM  $\beta$ -lactamases from clinical isolates of *Escherichia coli*. *Antimicrob. Agents Chemother.* 39:427-430.
- Imtiaz, U., E. Billings, J. R. Knox, E. K. Manavathu, S. A. Lerner, and S. Mobashery. 1993. Inactivation of class A  $\beta$ -lactamases by clavulanic acid: the role of arginine-244 in a proposed nonconcerted sequence of events. *J. Am. Chem. Soc.* 115:4435-4442.
- Imtiaz, U., E. K. Manavathu, S. Mobashery, and S. A. Lerner. 1994. Reversal of clavulanate resistance conferred by a Ser-244 mutant of TEM-1  $\beta$ -lactamase as a result of a second mutation (Arg to Ser at position 164) that enhances activity against ceftazidime. *Antimicrob. Agents Chemother.* 38:1134-1139.
- Knox, J. R. 1995. Extended-spectrum and inhibitor-resistant TEM-type  $\beta$ -lactamases: mutations, specificity, and three-dimensional structure. *Antimicrob. Agents Chemother.* 39:2593-2601.
- Kunkel, T. A., J. D. Roberts, and R. A. Zkour. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 154:367-382.
- Livermore, D. M. 1995.  $\beta$ -Lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* 8:557-584.
- Matagne, A., J. Lamotte-Brasseur, and J.-M. Frère. 1998. Catalytic properties of class A  $\beta$ -lactamases: efficiency and diversity. *Biochem. J.* 330:581-598.
- Nikaido, H. 1994. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* 264:382-388.
- Saves, I., O. Burlet-Schiltz, P. Swarén, F. Lefèvre, J.-M. Masson, J.-C. Promé, and J.-P. Samama. 1995. The asparagine to aspartic acid substitution at position 276 of TEM-35 and TEM-36 is involved in the  $\beta$ -lactamase resistance to clavulanic acid. *J. Biol. Chem.* 270:18240-18245.
- Sirot, D., C. Recule, E. B. Chaibi, L. Bret, J. Croize, C. Chanal-Claris, R. Labia, and J. Sirot. 1997. A complex mutant of TEM-1  $\beta$ -lactamase with mutations encountered in both IRT-4 and extended-spectrum TEM-15, produced by an *Escherichia coli* clinical isolate. *Antimicrob. Agents Chemother.* 41:1322-1325.
- Sowek, J. A., S. B. Singer, S. Ohringer, M. F. Malley, T. J. Dougherty, J. Z. Gougoutas, and K. Bush. 1991. Substitution of lysine at position 104 or 240 of TEM-1pTZ18R  $\beta$ -lactamase enhances the effect of serine-164 substitution on hydrolysis or affinity for cephalosporins and the monobactam aztreonam. *Biochemistry* 30:3179-3188.
- Spratt, B. G. 1994. Resistance to antibiotics mediated by target alterations. *Science* 265:388-393.
- Stapleton, P., K. Shannon, and I. Phillips. 1995. The ability of  $\beta$ -lactam antibiotics to select mutants with depressed  $\beta$ -lactamase synthesis from *Citrobacter freundii*. *J. Antimicrob. Chemother.* 36:483-496.
- Stapleton, P., P.-J. Wu, A. King, K. Shannon, G. French, and I. Phillips. 1995. Incidence and mechanisms of resistance to the combination of amoxicillin and clavulanic acid in *Escherichia coli*. *Antimicrob. Agents Chemother.* 39:2479-2483.
- Strynadka, N. C. J., H. Adachi, S. E. Jensen, K. Johns, A. Sielecki, C. Betzel, K. Sutoh, and M. N. G. James. 1992. Molecular structure of the acyl-enzyme intermediate in  $\beta$ -lactam hydrolysis at 1.7 Å resolution. *Nature* 359:700-705.
- Sutcliffe, J. G. 1978. Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. *Proc. Natl. Acad. Sci. USA* 75:3737-3741.
- Vakulenko, S. B., B. Geryk, L. P. Kotra, S. Mobashery, and S. A. Lerner.

1998. Selection and characterization of  $\beta$ -lactam- $\beta$ -lactamase inactivator-resistant mutants following PCR mutagenesis of the TEM-1  $\beta$ -lactamase gene. *Antimicrob. Agents Chemother.* **42**:1542-1548.
29. Vedel, G., A. Belaaouaj, L. Gilly, R. Labia, A. Philippon, P. N  vot, and G. Paul. 1992. Clinical isolates of *Escherichia coli* producing TRI  $\beta$ -lactamases: novel TEM-enzymes conferring resistance to  $\beta$ -lactamase inhibitors. *J. Antimicrob. Chemother.* **30**:449-462.
30. Zafaralla, G., E. K. Manavathu, S. A. Lerner, and S. Mobashery. 1992. Elucidation of the role of arginine-244 in the turnover processes of class A  $\beta$ -lactamases. *Biochemistry* **31**:3847-3852.
31. Zhou, X. Y., F. Bordon, D. Sirot, M.-D. Kitzis, and L. Gutmann. 1994. Emergence of clinical isolates of *Escherichia coli* producing TEM-1 derivatives or an OXA-1  $\beta$ -lactamase conferring resistance to  $\beta$ -lactamase inhibitors. *Antimicrob. Agents Chemother.* **38**:1085-1089.

## Codon usage tabulated from the GenBank genetic sequence data

Ken-nosuke Wada<sup>1</sup>, Shin-ichi Aota\*, Rye Tsuchiya, Fumie Ishibashi, Takashi Gojobori and Toshimichi Ikemura

National Institute of Genetics, Mishima, Shizuoka-ken 411 and <sup>1</sup>International Institute for Advanced Study of Social Information Science, Fujitsu Limited, Japan

In 1980 and 1981, Grantham and his colleague (1,2) reported the codon usages in a total of 161 protein genes in this journal, and in 1986 and 1988 we reported those in 1638 and 3681 genes, analyzing all the data available in those days (3,4). Now the codon usages in 11415 genes can be analyzed using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 62.0, Dec., 1989). Because of the growing size of the database, in this year a part of the data is listed. It is planned to distribute the electric version of the Sequence Supplement of this journal using a CD ROM, beginning in 1991. This year is a transition year, and thus we will send, upon request, a magnetic tape or a hard copy listing the codon usages in 11415 genes.

Table 1 lists the codon use in each of the 1543 nuclear genes registered in the GenBank Primate Sequence File. The LOCUS names given in the GenBank were used for designating individual genes, and the SHORT DIRECTORY of the GenBank is presented for defining each LOCUS name (Table 3).

To reveal the characteristics of the codon use of a wide range of organisms, as well as viruses and organella, the frequency (per one thousand) of codon use in each organism for which more than 20 genes are available was calculated by summing up numbers of codon use (Table 2). The number of genes summed for each organisms is given in the row designated as No. GENES, and the total codon number thus summed is given at the bottom row. Since the codon usage of each organism thus summed has been expressed in frequency per one thousand, it is easy to compare the codon-usage patterns among different organisms. Confirming the 'genome hypothesis' of Grantham et al. (1, 2), among taxonomically related organisms (e.g. among mammals) the codon-choice patterns resemble each other but they differ between distant organisms (see also scatter diagrams of Fig. 1). Synonymous codon-choice patterns in different genes of a single unicellular organism are known to be usually similar with each other regardless of gene functions and thus with the pattern listed in Table 2, (dialectal codon-choice pattern found for individual unicellular organisms, see ref. 5). However, codon-choice patterns in one higher vertebrate often differ significantly between different genes (5-7): The diverse codon-choice patterns found among genes of a single higher vertebrate have been pointed out in connection with the evident diversity in the G+C% at the codon third position among the genes (5,6). It should be stressed that the characteristic pattern for the mammals listed in Table 2 is obtained only after summing up the genes with varying

functions (3,4). When codon usages of approximately 10 or more genes with varying functions were summed up for each mammal, they usually resulted in a very similar pattern and thus in the pattern listed in Table 2, regardless of differences in the genes used for the summation (3,4). The fact that the pattern roughly common among the mammals does not depend on the genes used for the summation shows that this relates with general characteristics of their genomes: 1) deficiency of CpG and TpA dinucleotides (8); 2) paucity of genes in the A+T-rich genome portion, i.e. in A+T-rich isochores and in G/Q bands (see ref. 9, 10; thus C- and G-ending codons are preferred); 3) gross similarity of tRNA population between different organs of higher vertebrates (our unpublished data); 4) gross similarity of amino acid composition between different proteins, as well as between different mammals.

### METHODS

In selecting protein coding sequences we relied on the FEATURES tables of the GenBank, and only complete genes, starting with an initiation codon and ending with one of stop codons, were used in the analysis (see ref. 3 for details). In the GenBank, a group of consecutive genes whose entire region had been sequenced were registered under one LOCUS name. To distinguish the different genes belonging to a single LOCUS, symbol # followed by a number is added after the LOCUS name; the numbers represent the order of the peptides registered in the FEATURES of the GenBank. When introns of a gene have not been completely sequenced, some of its exons are registered in separate entries (LOCUS) in the GenBank. These exons belonging to the same gene but having different LOCUS names were combined, and the LOCUS name of the last exon followed by symbol \* was given to the gene thus combined (3,4). The order of the codons in the table is the same as the previous compilation (1-4).

### ACKNOWLEDGMENTS

The authors are very grateful to Dr. S. Osawa (Nagoya Univ.). This work was supported by grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan (No. 01656004).

\* Present address: Department of Health & Human Services, National Institute of Health, USA

## REFERENCES

1. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pav'e, A. (1980) *Nucl. Acids Res.* 8, r49-62.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nucl. Acids Res.* 9, r43-74.
3. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) *Nucl. Acids Res.* 14, r151-197.
4. Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) *Nucl. Acids Res.* 16, r315-402.
5. Ikemura, T. (1985) *Mol. Biol. Evol.* 2, 13-24.
6. Aota, S. and Ikemura, T. (1986) *Nucl. Acids Res.* 14, 6345-6355 & 8702.
7. Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F. (1988) *Nucl. Acids Res.* 16, 8207-8211.
8. Nussinov, R. (1984) *J. Mol. Evol.* 20 111-119.
9. Bernardi, G. and Bernardi, G. (1986) *Cold Spring Harbor Symp. Quant. Biol.* 51, 479-487.
10. Ikemura, T. and Aota, S. (1988) *J. Mol. Biol.* 203, 1-13.



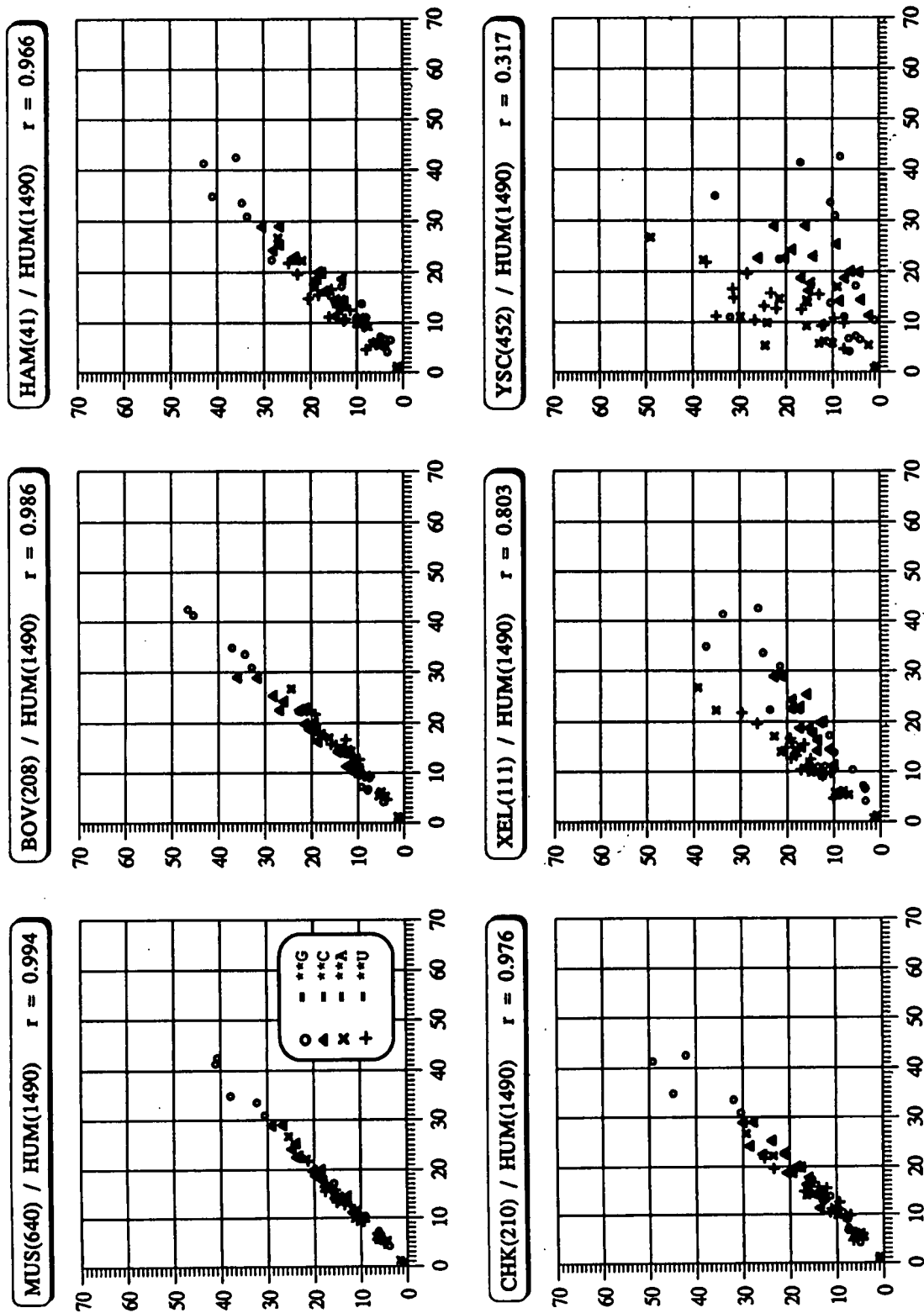


Fig. 1. Scatter diagrams listing a part of the data in Table 2. X-axis for human, Y-axis for individual organisms. The third letter of codons is specified using different symbols. (r=correlation coefficient)

**Table 1.** Codon usage in individual genes of primates (actual number of codons). Abbreviations for genes are defined in Table 3.

[illegible]

2	HUM HUM	
---	---	--

[illegible]

[illegible]

ARG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CGC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CGG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CGU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CGA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
AGG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81</																			

[illegible]





[illegible]

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	8																		

[illegible]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	5
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

[illegible]

[illegible]

UGA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
-----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

[illegible]

[illegible]











[illegible]



[illegible]





2d	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	11A	12A	13A	14A	15A	16A	17A	18A	19A	20A	21A	22A	23A	24A	25A	26A	27A	28A	29A	30A	31A	32A	33A	34A	35A	36A	37A	38A	39A	40A	41A	42A	43A	44A	45A	46A	47A	48A	49A	50A	51A	52A	53A	54A	55A	56A	57A	58A	59A	60A	61A	62A	63A	64A	65A	66A	67A	68A	69A	70A	71A	72A	73A	74A	75A	76A	77A	78A	79A	80A	81A	82A	83A	84A	85A	86A	87A	88A	89A	90A	91A	92A	93A	94A	95A	96A	97A	98A	99A	100A	101A	102A	103A	104A	105A	106A	107A	108A	109A	110A	111A	112A	113A	114A	115A	116A	117A	118A	119A	120A	121A	122A	123A	124A	125A	126A	127A	128A	129A	130A	131A	132A	133A	134A	135A	136A	137A	138A	139A	140A	141A	142A	143A	144A	145A	146A	147A	148A	149A	150A	151A	152A	153A	154A	155A	156A	157A	158A	159A	160A	161A	162A	163A	164A	165A	166A	167A	168A	169A	170A	171A	172A	173A	174A	175A	176A	177A	178A	179A	180A	181A	182A	183A	184A	185A	186A	187A	188A	189A	190A	191A	192A	193A	194A	195A	196A	197A	198A	199A	200A	201A	202A	203A	204A	205A	206A	207A	208A	209A	210A	211A	212A	213A	214A	215A	216A	217A	218A	219A	220A	221A	222A	223A	224A	225A	226A	227A	228A	229A	230A	231A	232A	233A	234A	235A	236A	237A	238A	239A	240A	241A	242A	243A	244A	245A	246A	247A	248A	249A	250A	251A	252A	253A	254A	255A	256A	257A	258A	259A	260A	261A	262A	263A	264A	265A	266A	267A	268A	269A	270A	271A	272A	273A	274A	275A	276A	277A	278A	279A	280A	281A	282A	283A	284A	285A	286A	287A	288A	289A	290A	291A	292A	293A	294A	295A	296A	297A	298A	299A	300A	301A	302A	303A	304A	305A	306A	307A	308A	309A	310A	311A	312A	313A	314A	315A	316A	317A	318A	319A	320A	321A	322A	323A	324A	325A	326A	327A	328A	329A	330A	331A	332A	333A	334A	335A	336A	337A	338A	339A	340A	341A	342A	343A	344A	345A	346A	347A	348A	349A	350A	351A	352A	353A	354A	355A	356A	357A	358A	359A	360A	361A	362A	363A	364A	365A	366A	367A	368A	369A	370A	371A	372A	373A	374A	375A	376A	377A	378A	379A	380A	381A	382A	383A	384A	385A	386A	387A	388A	389A	390A	391A	392A	393A	394A	395A	396A	397A	398A	399A	400A	401A	402A	403A	404A	405A	406A	407A	408A	409A	410A	411A	412A	413A	414A	415A	416A	417A	418A	419A	420A	421A	422A	423A	424A	425A	426A	427A	428A	429A	430A	431A	432A	433A	434A	435A	436A	437A	438A	439A	440A	441A	442A	443A	444A	445A	446A	447A	448A	449A	450A	451A	452A	453A	454A	455A	456A	457A	458A	459A	460A	461A	462A	463A	464A	465A	466A	467A	468A	469A	470A	471A	472A	473A	474A	475A	476A	477A	478A	479A	480A	481A	482A	483A	484A	485A	486A	487A	488A	489A	490A	491A	492A	493A	494A	495A	496A	497A	498A	499A	500A	501A	502A	503A	504A	505A	506A	507A	508A	509A	510A	511A	512A	513A	514A	515A	516A	517A	518A	519A	520A	521A	522A	523A	524A	525A	526A	527A	528A	529A	530A	531A	532A	533A	534A	535A	536A	537A	538A	539A	540A	541A	542A	543A	544A	545A	546A	547A	548A	549A	550A	551A	552A	553A	554A	555A	556A	557A	558A	559A	560A	561A	562A	563A	564A	565A	566A	567A	568A	569A	570A	571A	572A	573A	574A	575A	576A	577A	578A	579A	580A	581A	582A	583A	584A	585A	586A	587A	588A	589A	590A	591A	592A	593A	594A	595A	596A	597A	598A	599A	600A	601A	602A	603A	604A	605A	606A	607A	608A	609A	610A	611A	612A	613A	614A	615A	616A	617A	618A	619A	620A	621A	622A	623A	624A	625A	626A	627A	628A	629A	630A	631A	632A	633A	634A	635A	636A	637A	638A	639A	640A	641A	642A	643A	644A	645A	646A	647A	648A	649A	650A	651A	652A	653A	654A	655A	656A	657A	658A	659A	660A	661A	662A	663A	664A	665A	666A	667A	668A	669A	670A	671A	672A	673A	674A	675A	676A	677A	678A	679A	680A	681A	682A	683A	684A	685A	686A	687A	688A	689A	690A	691A	692A	693A	694A	695A	696A	697A	698A	699A	700A	701A	702A	703A	704A	705A	706A	707A	708A	709A	710A	711A	712A	713A	714A	715A	716A	717A	718A	719A	720A	721A	722A	723A	724A	725A	726A	727A	728A	729A	730A	731A	732A	733A	734A	735A	736A	737A	738A	739A	740A	741A	742A	743A	744A	745A	746A	747A	748A	749A	750A	751A	752A	753A	754A	755A	756A	757A	758A	759A	760A	761A	762A	763A	764A	765A	766A	767A	768A	769A	770A	771A	772A	773A	774A	775A	776A	777A	778A	779A	780A	781A	782A	783A	784A	785A	786A	787A	788A	789A	790A	791A	792A	793A	794A	795A	796A	797A	798A	799A	800A	801A	802A	803A	804A	805A	806A	807A	808A	809A	810A	811A	812A	813A	814A	815A	816A	817A	818A	819A	820A	821A	822A	823A	824A	825A	826A	827A	828A	829A	830A	831A	832A	833A	834A	835A	836A	837A	838A	839A	840A	841A	842A	843A	844A	845A	846A	847A	848A	849A	850A	851A	852A	853A	854A	855A	856A	857A	858A	859A	860A	861A	862A	863A	864A	865A	866A	867A	868A	869A	870A	871A	872A	873A	874A	875A	876A	877A	878A	879A	880A	881A	882A	883A	884A	885A	886A	887A	888A	889A	890A	891A	892A	893A	894A	895A	896A	897A	898A	899A	900A	901A	902A	903A	904A	905A	906A	907A	908A	909A	910A	911A	912A	913A	914A	915A	916A	917A	918A	919A	920A	921A	922A	923A	924A	925A	926A	927A	928A	929A	930A	931A	932A	933A	934A	935A	936A	937A	938A	939A	940A	941A	942A	943A	944A	945A	946A	947A	948A	949A	950A	951A	952A	953A	954A	955A	956A	957A	958A	959A	960A	961A	962A	963A	964A	965A	966A	967A	968A	969A	970A	971A	972A	973A	974A	975A	976A	977A	978A	979A	980A	981A	982A	983A	984A	985A	986A	987A	988A	989A	990A	991A	992A	993A	994A	995A	996A	997A	998A	999A	1000A
----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------

[illegible]

[illegible]

[illegible]

28	1A	1B	1C	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	2K	2L	2M	2N	2O	2P	2Q	2R	2S	2T	2U	2V	2W	2X	2Y	2Z	3A	3B	3C	3D	3E	3F	3G	3H	3I	3J	3K	3L	3M	3N	3O	3P	3Q	3R	3S	3T	3U	3V	3W	3X	3Y	3Z	4A	4B	4C	4D	4E	4F	4G	4H	4I	4J	4K	4L	4M	4N	4O	4P	4Q	4R	4S	4T	4U	4V	4W	4X	4Y	4Z	5A	5B	5C	5D	5E	5F	5G	5H	5I	5J	5K	5L	5M	5N	5O	5P	5Q	5R	5S	5T	5U	5V	5W	5X	5Y	5Z	6A	6B	6C	6D	6E	6F	6G	6H	6I	6J	6K	6L	6M	6N	6O	6P	6Q	6R	6S	6T	6U	6V	6W	6X	6Y	6Z	7A	7B	7C	7D	7E	7F	7G	7H	7I	7J	7K	7L	7M	7N	7O	7P	7Q	7R	7S	7T	7U	7V	7W	7X	7Y	7Z	8A	8B	8C	8D	8E	8F	8G	8H	8I	8J	8K	8L	8M	8N	8O	8P	8Q	8R	8S	8T	8U	8V	8W	8X	8Y	8Z	9A	9B	9C	9D	9E	9F	9G	9H	9I	9J	9K	9L	9M	9N	9O	9P	9Q	9R	9S	9T	9U	9V	9W	9X	9Y	9Z	10A	10B	10C	10D	10E	10F	10G	10H	10I	10J	10K	10L	10M	10N	10O	10P	10Q	10R	10S	10T	10U	10V	10W	10X	10Y	10Z	11A	11B	11C	11D	11E	11F	11G	11H	11I	11J	11K	11L	11M	11N	11O	11P	11Q	11R	11S	11T	11U	11V	11W	11X	11Y	11Z	12A	12B	12C	12D	12E	12F	12G	12H	12I	12J	12K	12L	12M	12N	12O	12P	12Q	12R	12S	12T	12U	12V	12W	12X	12Y	12Z	13A	13B	13C	13D	13E	13F	13G	13H	13I	13J	13K	13L	13M	13N	13O	13P	13Q	13R	13S	13T	13U	13V	13W	13X	13Y	13Z	14A	14B	14C	14D	14E	14F	14G	14H	14I	14J	14K	14L	14M	14N	14O	14P	14Q	14R	14S	14T	14U	14V	14W	14X	14Y	14Z	15A	15B	15C	15D	15E	15F	15G	15H	15I	15J	15K	15L	15M	15N	15O	15P	15Q	15R	15S	15T	15U	15V	15W	15X	15Y	15Z	16A	16B	16C	16D	16E	16F	16G	16H	16I	16J	16K	16L	16M	16N	16O	16P	16Q	16R	16S	16T	16U	16V	16W	16X	16Y	16Z	17A	17B	17C	17D	17E	17F	17G	17H	17I	17J	17K	17L	17M	17N	17O	17P	17Q	17R	17S	17T	17U	17V	17W	17X	17Y	17Z	18A	18B	18C	18D	18E	18F	18G	18H	18I	18J	18K	18L	18M	18N	18O	18P	18Q	18R	18S	18T	18U	18V	18W	18X	18Y	18Z	19A	19B	19C	19D	19E	19F	19G	19H	19I	19J	19K	19L	19M	19N	19O	19P	19Q	19R	19S	19T	19U	19V	19W	19X	19Y	19Z	20A	20B	20C	20D	20E	20F	20G	20H	20I	20J	20K	20L	20M	20N	20O	20P	20Q	20R	20S	20T	20U	20V	20W	20X	20Y	20Z	21A	21B	21C	21D	21E	21F	21G	21H	21I	21J	21K	21L	21M	21N	21O	21P	21Q	21R	21S	21T	21U	21V	21W	21X	21Y	21Z	22A	22B	22C	22D	22E	22F	22G	22H	22I	22J	22K	22L	22M	2
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---





[illegible]













[illegible]

HUMHCKB	HUMAN HEMOPOIETIC CELL PROTEIN-TYROSINE KINASE (HCK) GENE.	HUMHFN3	HUMAN INTERFERON-INDUCIBLE MRNA FRAGMENT (CONA 6-16).
HUMHCCR	HUMAN MRNA FOR PROTEIN HC (ALPHA-1-MICROGLOBULIN).	HUMHFN4	HUMAN GAMMA-INTERFERON-INDUCIBLE EARLY-RESPONSE GENE IP-10.
HUMHELA	HUMAN PAPILLOMAVIRUS (HPV) TYPE 18 ENCODING PROTEINS E6 AND	HUMHFN5	HUMAN INTERFERON-GAMMA RECEPTOR MRNA, COMPLETE CDS.
HUMHELB	HUMAN PAPILLOMAVIRUS (HPV) TYPE 18 PROTEINS E6, E7,	HUMHFN6	HUMAN 40-KDA KERATIN INTERMEDIATE FILAMENT PRECURSOR GENE.
HUMHER2A	HUMAN TYROSINE KINASE-TYPE RECEPTOR (HER2) MRNA, COMPLETE	HUMHGF1	HUMAN INSULIN-LIKE GROWTH FACTOR MRNA, COMPLETE CDS.
HUMHEXB13	HUMAN BETA-HEXOSAMINIDASE BETA-SUBUNIT (HEXB) GENE, EXONS 13	HUMHGF1B	HUMAN IGF-II GENE EXON 3 FOR INSULIN-LIKE GROWTH FACTOR II
HUMHF10	HUMAN HF-10 GENE MRNA.	HUMHGF1R	HUMAN MRNA FOR INSULIN-LIKE GROWTH FACTOR II RECEPTOR
HUMHFSP	HUMAN HANUKAH FACTOR SERINE PROTEASE (HUMF) MRNA, COMPLETE	HUMHGGFCRA	HUMAN LOW-AFFINITY IGG FC RECEPTOR (ALPHA-FC-GAMMA-RI)
HUMHGF	HUMAN HEPATOCTYTE GROWTH FACTOR (HGF) MRNA, COMPLETE CDS.	HUMHGGFCRB	HUMAN LOW-AFFINITY IGG FC RECEPTOR (BETA-FC-GAMMA-RI) MRNA, 1416BP
HUMHNGE	HUMAN MRNA FOR MITOCHONDRIAL HINGE PROTEIN.	HUMHGLAA	HUMAN LOW-AFFINITY IGG RECEPTOR CD16 (FCGR1B) MRNA.
HUMHS10G	HUMAN GENE FOR HISTONE H1(0).	HUMHGA2	HUMAN IG ACTIVE HEAVY CHAIN EPSILON-1 GENE, CONSTANT REGION.
HUMHS4	HUMAN H4 HISTONE GENE, COMPLETE CDS.	HUMHGBD	HUMAN UNPRODUCTIVELY REARRANGED IG MU-CHAIN MRNA V-REGION
HUMHS2A	HUMAN HISTONE H2A GENE.	HUMHGBP1	HUMAN REARRANGED AND TRUNCATED IG GAMMA HEAVY CHAIN DISEASE
HUMHS2B	HUMAN HISTONE H2B GENE.	HUMHGLAM2	HUMAN IG REARRANGED LAMBDA-CHAIN MRNA VUC-REGION SUBGROUP
HUMHS2A	HUMAN HISTONE H2B GENE.	HUMHGLBV	HUMAN MRNA FOR IG LAMBDA L-CHAIN V(VI)-J-C.
HUMHS2B	HUMAN HISTONE H2B GENE.	HUMHHP	HUMAN PLACENTAL RIBONUCLEASE INHIBITOR MRNA, COMPLETE CDS.
HUMHS2C	HUMAN HISTONE H2B GENE.	HUMHSHKD	HUMAN INTERFERON-INDUCIBLE 84 KD PROTEIN MRNA, COMPLETE CDS.
HUMHKA1	HUMAN HISTONE H2B GENE.	HUMHFS6	HUMAN INTERFERON-INDUCIBLE GENE IFI-202 5' REGION (86 KDA)
HUMHKA2	HUMAN HISTONE H2B GENE.	HUMHBP	HUMAN GAMMA-INTERFERON-INDUCIBLE PROTEIN (IP-30) MRNA.
HUMHLA1EA	HUMAN HLA-E CLASS I MRNA.	HUMHL1	HUMAN MONOCYTE INTERLEUKIN 1 (IL-1) MRNA, COMPLETE CDS.
HUMHLA13	HUMAN MHC CLASS I HLA-B13 GENE, COMPLETE CDS.	HUMHL1AA	HUMAN INTERLEUKIN 1 ALPHA MRNA, COMPLETE CDS.
HUMHLA8D	HUMAN MHC CLASS I HLA-B8 CHAIN GENE (A1 2; B5, 6), COMPLETE	HUMHL1AG	HUMAN GENE FOR INTERLEUKIN 1 ALPHA (IL-1 ALPHA).
HUMHLA8A	HUMAN MHC CLASS I HLA-B8A7 GENE, COMPLETE CDS.	HUMHL1B	HUMAN INTERLEUKIN-1 BETA GENE, COMPLETE CDS.
HUMHLADPA	HUMAN HLA DPA BETA-CHAIN MRNA, COMPLETE CDS.	HUMHL1BA	HUMAN INTERLEUKIN 1 BETA MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN HLA-DP-BETA 1 GENE AND HLA-DP-ALPHA-1 GENE EXON 1.	HUMHL1BR	HUMAN MRNA FOR INTERLEUKIN 1 BETA.
HUMHLADZA	HUMAN HLA CLASS II ALPHA CHAIN GENE DZ-ALPHA.	HUMHL1BX	HUMAN INTERLEUKIN 1 BETA GENE.
HUMHLA2	HUMAN CLASS II MHC GENE HLA-DZ7K EXONS 4-8 (BRUG CELL LINE).	HUMHL1P	HUMAN INTERLEUKIN-1 (IL-1) MRNA, COMPLETE CDS.
HUMHLA2A	HUMAN CLASS II MHC GENE HLA-DZ7K EXONS 4-8 (BRUG CELL LINE).	HUMHL2	HUMAN INTERLEUKIN 2 (IL-2) GENE, COMPLETE CODING SEQUENCE.
HUMHLA2B	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DO BETA CHAIN.	HUMHL2AB	HUMAN INTERLEUKIN 2 (IL-2) GENE, COMPLETE CDS, AND FLANKING
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2B	HUMAN INTERLEUKIN 2 GENE, CLONE PATTAC-3C277, COMPLETE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2B	HUMAN INTERLEUKIN 2 (IL-2) GENE, COMPLETE.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2BA	HUMAN INTERLEUKIN-3 RECEPTOR MRNA (LONG FORM), COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2BC	HUMAN INTERLEUKIN 2 RECEPTOR BETA CHAIN (P70-75) MRNA.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2BC	HUMAN INTERLEUKIN 2 RECEPTOR GENE, EXON 4.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL2B3	HUMAN INTERLEUKIN 2 (L2) GENE, EXON 4.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL3	HUMAN INTERLEUKIN 3 (IL-3) MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL3A	HUMAN INTERLEUKIN 3 (IL-3) MRNA, COMPLETE CDS, CLONE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL4	HUMAN INTERLEUKIN 4 (IL-4) MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL5	HUMAN INTERLEUKIN 5 (IL-5) GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL5A	HUMAN EOSINOPHIL DIFFERENTIATION FACTOR (INTERLEUKIN 5)
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL5R	HUMAN MRNA FOR T-CELL REPLACING FACTOR (INTERLEUKIN-5)
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL6	HUMAN INTERLEUKIN 6 RECEPTOR MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHL7A	HUMAN INTERLEUKIN 7 (IL-7) MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHLB2R	HUMAN MRNA FOR INTERLEUKIN 8/9-2 (B-CELL DIFFERENTIATION
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHMP	HUMAN INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP) MRNA.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC	HUMAN CYSTEINE-PROTEINASE INHIBITOR (CST1) GENE, COMPLETE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC3	HUMAN CYSTEINE-PROTEINASE INHIBITOR (CST2) GENE, EXON 3.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC4	HUMAN INHIBIN A-SUBUNIT MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC42	HUMAN PREPROINHIBIN A GENE EXON 2 FRAGMENT MACROPHAGE CELL
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC5A	HUMAN OVARIAN BETA-A INHIBIN MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC6	HUMAN INTERFERON-INDUCIBLE GENE IFI-54K 5' FLANK.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC7	HUMAN INSULIN GENE INCLUDING 5' AND 3' FLANKS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC8	HUMAN ALPHA-TYPE INSULIN GENE AND 5' FLANKING POLYMORPHIC
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC9	HUMAN INSULIN RECEPTOR MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC10	HUMAN INSULIN RECEPTOR MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC11	HUMAN INT-1 MAMMARY ONCOGENE.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC12	HUMAN INVOLUCRIN GENE, EXON 2.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC13	HUMAN INTERPHOTORECEPTOR RETINOID-BINDING PROTEIN (IRBP)
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC14	HUMAN INSULIN-RESPONSIVE GLUCOCORTICOID TRANSPORTER (GLUT4) MRNA.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC15	HUMAN IGF-54K GENE (INTERFERON STIMULATED GENE) ENCODING A
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC16	HUMAN INTERFERON-INDUCED 18-KD PROTEIN (IRG) GENE, EXON 2.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC17	HUMAN BK PROTEIN (EXHIBITING A SLOWLY ACTIVATING CHANNEL
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC18	HUMAN MRNA FOR SECOND PROTEIN OF INTER-ALPHA-TRYPTICIN
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC19	HUMAN MRNA FOR LYMPHOCYTE IGE RECEPTOR (LOW AFFINITY
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC20	HUMAN C-JUN PROTO ONCOGENE ENCODING JUN, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC21	HUMAN BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE (E2)
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC22	HUMAN GLANDULAR KALLIKREIN GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC23	HUMAN KERATIN 18 (K18) GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC24	HUMAN MRNA FOR KERATIN 18.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC25	HUMAN KERATIN TYPE 8 (K8) MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC26	HUMAN DNA FOR 65KD KERATIN TYPE 8 EXONS 8,7,6,5 AND
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC27	HUMAN EPIDERMAL 67-KDA KERATIN GENE, EXONS 7, 8, AND 9.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC28	HUMAN CYTOKERATIN 8 MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC29	HUMAN MRNA FOR CYTOKERATIN 15.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC30	HUMAN 50 KD TYPE I EPIDERMAL KERATIN GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC31	HUMAN K68 (EPIDERMAL KERATIN, TYPE 8) GENE, EXON 9.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC32	HUMAN POT. PSEUDO-KERATIN K18 TYPE 1, EXON 4.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC33	HUMAN KINNOGEN GENE, EXON 10, ENCODING BRADYKININ AND EXON
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC34	HUMAN ACIDIC KERATIN-10 MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC35	HUMAN KU AUTOIMMUNE ANTIGEN GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC36	HUMAN ALPHA-LACTALBUMIN GENE, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC37	HUMAN PRE-ALPHA-LACTALBUMIN MRNA.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC38	HUMAN LAMININ B1 CHAIN MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC39	HUMAN MRNA FOR NUCLEAR ENVELOPE PROTEIN LAMIN A PRECURSOR.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC40	HUMAN COLIN CARCINOMA LAMININ-BINDING PROTEIN MRNA, COMPLETE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC41	HUMAN LAMININ B2 CHAIN MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC42	HUMAN LAMIN C MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC43	HUMAN MRNA FOR NUCLEAR ENVELOPE PROTEIN LAMIN C PRECURSOR.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC44	HUMAN LAMP-1 MRNA ENCODING LYSO-SOMAL MEMBRANE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC45	HUMAN LAMP-2 GENE ENCODING LYSO-SOMAL MEMBRANE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC46	HUMAN LEUKOCYTE ADHESION RECEPTOR (LFA-1/MAAC-1/150,85
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC47	HUMAN LEUKOCYTE ADHESION RECEPTOR ALPHA SUBUNIT MRNA.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC48	HUMAN LIPOCORTIN-H MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC49	HUMAN LIPOCORTIN-V MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC50	HUMAN MRNA FOR T200 LEUKOCYTE COMMON ANTIGEN (CD45, LC-A).
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC51	HUMAN MRNA FOR LEUKOCYTE COMMON ANTIGEN (T200).
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC52	HUMAN LECTIN-CHOLESTEROL ACYLTRANSFERASE MRNA, COMPLETE
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC53	HUMAN GENE FOR LECTIN-CHOLESTEROL ACYLTRANSFERASE (LCAAT).
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC54	HUMAN LYMPHOCYTE CLATHRIN LIGHT CHAIN A MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC55	HUMAN LYMPHOCYTE CLATHRIN LIGHT CHAIN B MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC56	HUMAN TONILLAR LYMPHOCYTE L278 MRNA INDUCED BY TPA OR PHA
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC57	HUMAN LACTATE DEHYDROGENASE-A ISOZYME MRNA, COMPLETE CDS.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC58	HUMAN LACTATE DEHYDROGENASE-B GENE, EXON 7.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC59	HUMAN LACTATE DEHYDROGENASE-B GENE EXON 8 (EC 1.1.1.27).
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC60	HUMAN MRNA FOR LACTATE DEHYDROGENASE-B (LDH-B).
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC61	HUMAN TESTIS-SPECIFIC LACTATE DEHYDROGENASE (LDHC4, LDH0)
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC62	HUMAN MRNA FOR APOLOPROTEIN B-100.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC63	HUMAN LOW DENSITY LIPOPROTEIN RECEPTOR GENE, EXON 18.
HUMHLADPB	HUMAN MRNA FOR HLA-D CLASS II ANTIGEN DPM2 BETA CHAIN.	HUMHNC64	HUMAN MRNA FOR LDL-RECEPTOR RELATED PROTEIN.







HUMPA2A8	HUMAN PLASMINOGEN ACTIVATOR INHIBITOR-2 (PAI-2) GENE, EXON	HUMPR0F	HUMAN PROFILIN MRNA, COMPLETE CDS.
HUMPA2B	HUMAN PLASMINOGEN ACTIVATOR INHIBITOR 2 (PAI-2) MRNA,	HUMPR0LA	HUMAN CATHEPSIN L GENE, COMPLETE CDS.
HUMPA2R	HUMAN MRNA FOR APO-BERPIN (PLASMINOGEN ACTIVATOR INHIBITOR	HUMPROLEU	HUMAN SECRETORY GRANULE PROTEOGLYCAN PEPTIDE CORE MRNA,
HUMPA4A	HUMAN PLASMINOGEN ACTIVATOR INHIBITOR-1 GENE, EXONS 2-6	HUMPROT1	HUMAN TESTIS SPECIFIC MRNA FOR PROTAMINE 1 (P1).
HUMPA4R	HUMAN MRNA FOR ENDOTHELIAL PLASMINOGEN ACTIVATOR INHIBITOR	HUMPROTP	HUMAN ENDOCYTOSIS PROTON PUMP SUBUNIT MRNA, COMPLETE CDS.
HUMPALA	HUMAN PREALBUMIN MRNA, COMPLETE CDS.	HUMPRP	HUMAN PRION PROTEIN (PRP) MRNA, COMPLETE CDS.
HUMPALB	HUMAN PREALBUMIN MRNA, COMPLETE CDS.	HUMPRP2A	HUMAN PRION PROTEIN 27-30 MRNA, COMPLETE CDS.
HUMPALC	HUMAN SERUM PREALBUMIN GENE.	HUMPRPA	HUMAN PRP2 LOCUS SALIVARY PROLINE-RICH PROTEIN MRNA (PR1
HUMPALD	HUMAN PREALBUMIN GENE, COMPLETE CDS.	HUMPRPB	HUMAN PRP1 LOCUS SALIVARY PROLINE-RICH PROTEIN MRNA (PF
HUMPALF3	HUMAN MUTANT PREALBUMIN GENE DIRECTLY LINKED TO FAMILIAL	HUMPRPC	HUMAN PRP1 LOCUS SALIVARY PROLINE-RICH PROTEIN MRNA, CLONE
HUMPALFAP	HUMAN PREALBUMIN MRNA IN INDIVIDUALS WITH FAMILIAL	HUMPRPD	HUMAN PRP1 LOCUS SALIVARY PROLINE-RICH PROTEIN MRNA, CLONE
HUMPAP1	HUMAN PLACENTA ANTICOAGULANT PROTEIN PP4 MRNA, COMPLETE CDS.	HUMPRPE	HUMAN PRP4 LOCUS SALIVARY PROLINE-RICH PROTEIN MRNA, COMPLETE
HUMPAP2	HUMAN PROTEIN PP4-X MRNA, COMPLETE CDS.	HUMPRPF	HUMAN PRP1 GENE (HAEI-TYPE SUBFAMILY; ALLELE PRP1-4)
HUMPA2D	HUMAN GLYCOPHORIN C (P85-2) MRNA, COMPLETE CDS.	HUMPRPH1	HUMAN PRP2 GENE (HAEI-TYPE SUBFAMILY; ALLELE PRP2-1)
HUMPB0D1	HUMAN MRNA FOR PORPHYRINOGEN DEAMINASE (PBG-D, EC	HUMPRPH2	HUMAN MRNA FOR PHOSPHORIBOSYL PYRIMIDINATE SYNTHETASE
HUMPB0D2	HUMAN MRNA FOR NON-ERYTHROPOIETIC PORPHYRINOGEN DEAMINASE	HUMPRP2	HUMAN P2 MRNA, COMPLETE CDS, FROM BREAST BREAST CANCER CELL
HUMPCAR	HUMAN MRNA FOR CALCIUM DEPENDENT PROTEASE (SMALL SUBUNIT)	HUMPS2	HUMAN ESTROGEN-RESPONSIVE GENE P22 EXON 3.
HUMPCAR	HUMAN PROLIFERATING CELL NUCLEAR ANTIGEN GENE, COMPLETE CDS.	HUMPS23	HUMAN PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN/FETAL LIVER
HUMPCG7	HUMAN PLATELET-DERIVED GROWTH FACTOR A-CHAIN GENE, EXON 7.	HUMPSG1A	HUMAN PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN/FETAL LIVER
HUMPCG7A	HUMAN PLATELET-DERIVED GROWTH FACTOR (PDGF) A CHAIN GENE,	HUMPSG2A	HUMAN PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN/FETAL LIVER
HUMPCG7B	HUMAN PLATELET-DERIVED GROWTH FACTOR A-CHAIN	HUMPSG3A	HUMAN PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN/FETAL LIVER
HUMPCG7C	HUMAN PLATELET-DERIVED GROWTH FACTOR A-CHAIN	HUMPSPA	HUMAN PULMONARY SURFACTANT-ASSOCIATED PROTEIN MRNA, COMPLETE
HUMPCG7D	HUMAN PLATELET-DERIVED GROWTH FACTOR BETA (PDGF-B) MRNA,	HUMPSPB	HUMAN PULMONARY SURFACTANT-ASSOCIATED PROTEIN PBP-8 MRNA,
HUMPCG7E	HUMAN PLATELET-DERIVED GROWTH FACTOR (PDGF) RECEPTOR MRNA,	HUMPSPB2	HUMAN PULMONARY SURFACTANT-ASSOCIATED PROTEIN PBP-8 MRNA,
HUMPCG7F	HUMAN PLATELET-DERIVED GROWTH FACTOR (PDGF) RECEPTOR MRNA,	HUMPSPT	HUMAN PANCREATIC SECRETORY TRYPSIN INHIBITOR (PSTI) MRNA,
HUMPCG7G	HUMAN PYRUVATE DEHYDROGENASE ALPHA SUBUNIT MRNA, COMPLETE	HUMPSPTA	HUMAN PANCREATIC SECRETORY TRYPSIN INHIBITOR (PSTI) GENE,
HUMPCG7H	HUMAN PYRUVATE DEHYDROGENASE E1 ALPHA (PDH-E1A) SUBUNIT GENE,	HUMPSPTB	HUMAN PANCREATIC SECRETORY TRYPSIN INHIBITOR (PSTI) GENE,
HUMPCG7I	HUMAN PYRUVATE DEHYDROGENASE BETA SUBUNIT MRNA, COMPLETE	HUMPSPTC	HUMAN PROTHROMBIN ALPHA MRNA (PROT-ALPHA), COMPLETE CDS.
HUMPCG7J	HUMAN PEPISINNOGEN GENE, EXON 8.	HUMPSPTD	HUMAN PARATHYROID (PTH) GENE, CODING REGION AND 3' FLANK,
HUMPCG7K	HUMAN PEPISINNOGEN A (150) GENE, EXON 8.	HUMPSPTL	HUMAN PARATHYROID-LIKE PROTEIN (ASSOCIATED WITH HUMORAL
HUMPCG7L	HUMAN PEPISINNOGEN C GENE, EXON 8.	HUMPSPTL2	HUMAN PARATHYROID HORMONE-LIKE PROTEIN (PLH) GENE, EXON 4.
HUMPCG7M	HUMAN PEPISINNOGEN GENE, EXON 8.	HUMPSPTL3	HUMAN RENAL CARCINOMA PARATHYROID HORMONE-LIKE PEPTIDE MRNA,
HUMPCG7N	HUMAN PROLIDASE (MDCOPEPTIDASE) MRNA, COMPLETE CDS.	HUMPSPTL4	HUMAN PARATHYROID HORMONE-RELATED PROTEIN MRNA, COMPLETE
HUMPCG7O	HUMAN PERFORMIN MRNA, COMPLETE CDS.	HUMPSPTL5	HUMAN PUMP-1 MRNA HOMOLOG TO METALLOPROTEINASE, COLLAGENASE
HUMPCG7P	HUMAN MRNA FOR GLUTATHIONE PEROXIDASE.	HUMPSPTL6	HUMAN PYRUVATE DEHYDROGENASE ALPHA SUBUNIT MRNA, COMPLETE
HUMPCG7Q	HUMAN MUSCLE PHOSPHOFRUCTOKINASE (PFKM) MRNA, COMPLETE CDS.	HUMPSPTL7	HUMAN PROLYL 4-HYDROXYLASE ALPHA SUBUNIT MRNA, COMPLETE CDS.
HUMPCG7R	HUMAN PHOSPHOGLYCERATE MUTASE MUSCLE-SPECIFIC SUBUNIT MRNA,	HUMPSPTL8	HUMAN PROLYL 4-HYDROXYLASE ALPHA SUBUNIT MRNA, COMPLETE CDS.
HUMPCG7S	HUMAN PHOSPHOGLYCERATE MUTASE (PGAM-M) GENE, COMPLETE CDS.	HUMPSPTL9	HUMAN MITOCHONDRIAL UBIQUINONE-BINDING PROTEIN MRNA, COMPLETE
HUMPCG7T	HUMAN PHOSPHOGLYCERATE KINASE (PGK) MRNA, EXONS 2 TO LAST.	HUMPSPTL10	HUMAN RAS2 MRNA, YPT1-RELATED AND MEMBER OF RAS FAMILY.
HUMPCG7U	HUMAN TESTIS-SPECIFIC PGK-2 GENE FOR PHOSPHOGLYCERATE KINASE	HUMPSPTL11	HUMAN MRNA FOR RAF ONCOGENE.
HUMPCG7V	HUMAN X-LINKED PHOSPHOGLYCERATE KINASE GENE, EXON 11.	HUMPSPTL12	HUMAN CELLULAR RETINALDEHYDE-BINDING PROTEIN MRNA, COMPLETE
HUMPCG7W	HUMAN MRNA FOR PROTEIN GENE PRODUCT (PGP) S.S.	HUMPSPTL13	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG7X	HUMAN PROGESTERONE RECEPTOR MRNA, COMPLETE CDS.	HUMPSPTL14	HUMAN MRNA FOR RETINOIC ACID RECEPTOR.
HUMPCG7Y	HUMAN MRNA FOR PLASMA GELSOLIN.	HUMPSPTL15	HUMAN RETINOIC ACID RECEPTOR-EPSILON MRNA, COMPLETE CDS.
HUMPCG7Z	HUMAN PHENYLALANINE HYDROXYLASE MRNA, COMPLETE CDS.	HUMPSPTL16	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8A	HUMAN PREPRO-INSULIN-LIKE GROWTH FACTOR II (PREPRO-IGF-II)	HUMPSPTL17	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8B	HUMAN PMS-1 ONCOGENE MRNA, COMPLETE CDS.	HUMPSPTL18	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8C	HUMAN PMS-1 PROTO-ONCOGENE GENE, COMPLETE CDS.	HUMPSPTL19	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8D	HUMAN PROLACTIN-INDUCIBLE PROTEIN (PIP) MRNA, COMPLETE CDS.	HUMPSPTL20	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8E	HUMAN MRNA FOR CAMP-DEPENDENT PROTEIN KINASE CATALYTIC	HUMPSPTL21	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8F	HUMAN BETA TYPE PROTEIN KINASE C MRNA, COMPLETE CDS.	HUMPSPTL22	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8G	HUMAN MRNA FOR PROTEIN KINASE C (PKC) TYPE BETA 1.	HUMPSPTL23	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8H	HUMAN MRNA FOR PROTEIN KINASE C (PKC) TYPE BETA 1.	HUMPSPTL24	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8I	HUMAN PYRUVATE KINASE TYPE L MRNA, COMPLETE CDS.	HUMPSPTL25	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8J	HUMAN M2-TYPE PYRUVATE KINASE MRNA, COMPLETE CDS.	HUMPSPTL26	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8K	HUMAN PLACENTAL LACTOGEN HORMONE: HPL-3 GENE AND FLANKS.	HUMPSPTL27	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8L	HUMAN PANCREATIC PHOSPHOLIPASE A-2 (PLA-2) GENE, EXON 4.	HUMPSPTL28	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8M	HUMAN LUNG PHOSPHOLIPASE A-2 (PLA-2) MRNA, COMPLETE CDS.	HUMPSPTL29	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8N	HUMAN PLAKOGLOBIN MRNA, COMPLETE CDS.	HUMPSPTL30	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8O	HUMAN PLACENTAL ALKALINE PHOSPHATASE-LIKE GENE, COMPLETE	HUMPSPTL31	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8P	HUMAN L-PLASTIN POLYPEPTIDE MRNA, COMPLETE CDS, CLONE P4.	HUMPSPTL32	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8Q	HUMAN L-PLASTIN POLYPEPTIDE MRNA, COMPLETE CDS, CLONE P17.	HUMPSPTL33	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8R	HUMAN TISSUE-TYPE PLASMINOGEN ACTIVATOR (T-PA) MRNA,	HUMPSPTL34	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8S	HUMAN PLA-X MRNA.	HUMPSPTL35	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8T	HUMAN PLACENTAL LACTOGEN HORMONE: HPL-4 MRNA.	HUMPSPTL36	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8U	HUMAN PLATELET MEMBRANE GLYCOPROTEIN IIb MRNA, COMPLETE CDS.	HUMPSPTL37	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8V	HUMAN MYELIN PROTEOLIPID PROTEIN GENE, EXON 7.	HUMPSPTL38	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8W	HUMAN PROTEOLIPID PROTEIN MRNA (PLP), COMPLETE CDS.	HUMPSPTL39	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8X	HUMAN CMB MYELIN MAJOR PROTEOLIPID COMPONENT (DM-20) MRNA,	HUMPSPTL40	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8Y	HUMAN PULMONARY SURFACTANT PROTEIN C GENE, COMPLETE CDS.	HUMPSPTL41	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG8Z	HUMAN PLATELET MEMBRANE GLYCOPROTEIN IIIa BETA SUBUNIT MRNA,	HUMPSPTL42	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9A	HUMAN MRNA FOR PLASMINOGEN.	HUMPSPTL43	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9B	HUMAN PLASMA MEMBRANE CA2+ PUMPING ATPASE MRNA, COMPLETE	HUMPSPTL44	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9C	HUMAN PHENYLETHANOLAMINE N-METHYLTRANSFERASE MRNA, COMPLETE	HUMPSPTL45	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9D	HUMAN PHENYLETHANOLAMINE N-METHYLTRANSFERASE GENE, COMPLETE	HUMPSPTL46	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9E	HUMAN PURINE NUCLEOSIDE PHOSPHORYLASE (PNP) MRNA, COMPLETE	HUMPSPTL47	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9F	HUMAN PURINE NUCLEOSIDE PHOSPHORYLASE GENE, EXON 6.	HUMPSPTL48	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9G	HUMAN POLYMERASE BETA MRNA, COMPLETE CDS.	HUMPSPTL49	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9H	HUMAN DNA POLYMERASE ALPHA-SUBUNIT MRNA, COMPLETE CDS.	HUMPSPTL50	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9I	HUMAN POLY(ADP-RIBOSE) POLYMERASE MRNA, COMPLETE CDS.	HUMPSPTL51	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9J	HUMAN MRNA FOR POLYA BINDING PROTEIN.	HUMPSPTL52	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9K	HUMAN PROOPOMELANOCORTIN (POMC) GENE, COMPLETE CDS.	HUMPSPTL53	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9L	HUMAN PROOPOMELANOCORTIN (POMC) GENE, EXON 3.	HUMPSPTL54	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9M	HUMAN POLIOVIRUS RECEPTOR MRNA, CLONE H208.	HUMPSPTL55	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9N	HUMAN POLIOVIRUS RECEPTOR MRNA, CLONE H208.	HUMPSPTL56	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9O	HUMAN PLACENTAL PROTEIN (PP12) MRNA, COMPLETE CDS.	HUMPSPTL57	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9P	HUMAN INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN/PLACENTAL	HUMPSPTL58	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9Q	HUMAN PLACENTAL PROTEIN 14 (PP14) MRNA, COMPLETE CDS.	HUMPSPTL59	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9R	HUMAN GENE FOR PP15 (PLACENTAL PROTEIN 15).	HUMPSPTL60	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9S	HUMAN PROTEIN PHOSPHATASE 2A MRNA, COMPLETE CDS.	HUMPSPTL61	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9T	HUMAN ACIDIC RIBOSOMAL PHOSPHOPROTEIN P0 MRNA, COMPLETE CDS.	HUMPSPTL62	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9U	HUMAN ACIDIC RIBOSOMAL PHOSPHOPROTEIN P1 MRNA, COMPLETE CDS.	HUMPSPTL63	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9V	HUMAN ACIDIC RIBOSOMAL PHOSPHOPROTEIN P2 MRNA, COMPLETE CDS.	HUMPSPTL64	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9W	HUMAN PANCREATIC PROTEASE E MRNA, COMPLETE CDS.	HUMPSPTL65	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9X	HUMAN SHRN POLYPEPTIDE B MRNA, COMPLETE CDS.	HUMPSPTL66	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9Y	NUCLEOTIDE SEQUENCE OF THE CDNA INSERT OF LAMBDA PK129	HUMPSPTL67	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPCG9Z	HUMAN PANCREATIC POLYPEPTIDE (PP) AND ICOSAPETIDE PRECURSOR	HUMPSPTL68	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0A	HUMAN PANCREATIC POLYPEPTIDE GENE, COMPLETE CDS.	HUMPSPTL69	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0B	HUMAN PROTECTIVE PROTEIN MRNA, COMPLETE CDS.	HUMPSPTL70	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0C	HUMAN VITAMIN K-DEPENDENT PLASMA PROTEIN 8 MRNA, COMPLETE	HUMPSPTL71	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0D	HUMAN PANCREATIC POLYPEPTIDE Y MRNA, COMPLETE CDS.	HUMPSPTL72	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0E	HUMAN PREA4 GENE FOR ALZHEIMER'S DISEASE A4 AMYLOID PROTEIN	HUMPSPTL73	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0F	HUMAN PROTEIN C GENE, EXON 9 OF 9.	HUMPSPTL74	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0G	HUMAN PROTEIN C GENE, COMPLETE CDS.	HUMPSPTL75	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0H	HUMAN PROTEIN C MRNA, COMPLETE CDS.	HUMPSPTL76	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0I	HUMAN PRG GENE (ALT. HPG) ENCODING HEMOPHILIC PROTEOGLYCAN	HUMPSPTL77	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0J	HUMAN PREPROLACTIN (PRL) MRNA.	HUMPSPTL78	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0K	HUMAN PROLACTIN GENE, EXON 5.	HUMPSPTL79	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.
HUMPD0L	HUMAN PROLYL 4-HYDROXYLASE BETA-SUBUNIT AND DISULFIDE	HUMPSPTL80	HUMAN RAS2 MRNA FOR RAS-RELATED PROTEIN.



## FOR TABLE 2

HUM	HUMAN
HAM	HAMSTER
MUS	MOUSE
RAT	RAT
BOV	BOVINE
DOG	DOG
PIG	PIG
RAB	RABBIT
SHP	SHEEP
CHK	CHICKEN
FBS	FISH (GROUP B)
XEL	XENOPUS LAEVIS
BMO	BOMBYX MORI
CEL	CAENORHABDITIS ELEGANS
DOI	DICTYOSTELIUM DISCOIDEUM
DRO	DROSOPHILA MELANOGASTER
PFA	PLASMODIUM FALCIPARUM
SCM	SCHISTOSOMA MANSONI
SUS	SEA URCHIN
TRB	TRYPANOSOMA BRUCEI
ASN	ASPERGILLUS NIDULANS
ATH	ARABIDOPSIS THALIANA
BLY	BARLEY
MZE	MAIZE
NEU	NEUROSPORA
PEA	PEA
POT	POTATO
SOY	SOYBEAN
TOM	TOMATO
WHT	WHEAT
YSC	SACCHAROMYCES CEREVISIAE
YSK	YEAST KLUYVEROMYCES
YSP	SCHIZOSACCHAROMYCES POMBE
ANA	ANABENA
ATU	AGROBACTERIUM TUMEFACIENS
AVI	AZOTOBACTER VINELANDII
BME	BACILLUS MEGATERIUM
BST	BACILLUS STEAROTHERMOPHILUS
BSU	BACILLUS SUBTILIS
BTH	BACILLUS THURINGIENSIS
CHT	CHLAMYDIA TRACHOMATIS
CLO	CLOSTRIDIUM
ECO	ESCHERICHIA COLI
FDI	FREMYELLA DIPLOPHON
HAL	HALOBACTERIUM HALOBIVM
FPL	F PLASMID
KPN	KLEBSIELLA PNEUMONIA
PSE	PSEUDOMONAS
RCA	RHODOBACTER CAPSULATUS
RHL	RHIZOBIUM LEGUMINOSARUM
RHM	RHIZOBIUM MELILOTI
SSP	SULFOLOBUS VIRUS-LIKE PARTICLE
STA	STAPHYLOCOCCUS AUREUS
STM	STREPTOMYCES
STR	STREPTOCOCCUS
STY	SALMONELLA TYPHIMURIVM
SYN	SYNECHOCOCCUS
TIP	TI PLASMID
TRN	TRANSPOSON TN
VIB	VIBRIO
AD8	ADENOVIRUS TYPE 2
FLA	INFLUENZA A
FLB	INFLUENZA B
HIV	HUMAN IMMUNODEFICIENCY VIRUS
HPB	HEPATITIS B VIRUS
HS1	HERPES SIMPLEX VIRUS TYPE 1
HS4	EPSTEIN-BARR VIRUS
HS5	HUMAN CYTOMEGALOVIRUS
MCV	CUCUMBER MOSAIC VIRUS
NDV	NEWCASTLE DISEASE VIRUS
PIF	HUMAN PARAINFLUENZA VIRUS
PLY	POLYOMA VIRUS
RPH	HUMAN PAPILLOMAVIRUS
REO	REOVIRUS
SV	SIMIAN IMMUNODEFICIENCY VIRUS
VAC	VACCINIA VIRUS
VAZ	VARICELLA-ZOSTER VIRUS
VSV	VESICULAR STOMATITIS VIRUS
F1C	BACTERIOPHAGE F1
LAM	BACTERIOPHAGE LAMBDA
P22	BACTERIOPHAGE P22
PT4	BACTERIOPHAGE T4
PT7	BACTERIOPHAGE T7
PZA	BACTERIOPHAGE PZA
MPO CP	MARCHANTIA POLYMORPHA CHLOROPLAST
MZE CP	MAIZE CHLOROPLAST
PEA CP	PEA CHLOROPLAST
SPI CP	SPINACH CHLOROPLAST
TOS CP	TOBACCO CHLOROPLAST
YSC MT	S. CEREVISIAE MITOCHONDRIA

---

## Bioluminescent Click Beetles Revisited

---

Keith V. Wood\*, Y. Amy Lam and William D. McElroy

Department of Chemistry, M-001, University of California, San Diego, La Jolla, CA 92093, USA

Howard H. Seliger

Department of Biology, Johns Hopkins University, Baltimore, MD, USA

---

In studying beetle bioluminescence in the early 1960s, Dr McElroy and his colleagues found that the Jamaican click beetle, *Pyrophorus plagiophthalmus*, was capable of emitting different colours of light. They further found that the luciferin substrate used by this beetle was the same as that in the firefly, demonstrating that the different colours of bioluminescence were due to differences in the structure of the luciferases. We have recently cloned cDNAs from this beetle species which code for at least four different luciferases. The luciferases are distinguishable by their different colours of bioluminescence when expressed in *Escherichia coli*. The sequence differences between these different luciferases are few, so the amino acids responsible for the different colours of emission must also be few. Through the construction of hybrid luciferases, by rearranging fragments of the original cDNA clones, we have identified some of these amino acid determinants of colour.

**Keywords:** Firefly luciferase; click beetle luciferases; bioluminescence spectra

### INTRODUCTION

In 1963, a scientific expedition to Jamaica was led by William McElroy to study bioluminescence. There they encountered the 'kitty boo', a local name for the large bioluminescent click beetle, *Pyrophorus plagiophthalmus*. The beetle attracted the scientists' attention because of its ability to emit different colours of bioluminescence, a feature not found among species of true fireflies (Seliger *et al.*, 1964). It has two sets of light organs, a pair on the dorsal surface of the head, and a single light organ in a cleft on the ventral surface of the abdomen. The dorsal organs emit greenish light, while the ventral organ usually emits yellow or orange light. But the differences in colour do not occur only within individual specimens. Considerable variation also

occurs between specimens for each set of organs. The colour emitted from the dorsal organs can range from green to yellow-green, and that of the ventral can range from green to orange. It was shown by these scientists that the differences in colour were not due to differences in the substrates of the luminescent reactions. The luciferases of these click beetle, or any other bioluminescent beetle, utilise ATP and the same luciferin that was first elucidated in the chemistry of the firefly luciferase. It was concluded that the different colours of bioluminescence were caused by subtle differences in the interaction of the substrates with the enzyme. Unfortunately, attempts to study these luciferases further were limited by the difficulties of collecting sufficient quantities of the beetles. Recently, we have been able to overcome this problem with the use of

\*Author for correspondence.

molecular cloning techniques. cDNAs generated from the ventral light organ of this beetle have been cloned and expressed in *Escherichia coli* (Wood *et al.*, 1988a, 1988b). These cDNA clones are of four different types, distinguishable by the colour of light emitted by the luciferases they code: green (546 nm), yellow-green (560 nm), yellow (578 nm), and orange (593 nm). These cDNA clones not only accord a practical source of the click beetle luciferases, but also provided a means of separating the different type since each clone is expressed in a separate host. Given here is the first report on the structural features of these click beetle luciferases that are responsible for the colour of light emitted.

#### AMINO ACID SEQUENCES AND THE DETERMINANTS OF COLOUR

Four cDNA clones, each capable of eliciting one of the four colours of bioluminescence in *E. coli*, were sequenced. With this information the amino acid sequences of the corresponding luciferases were deduced (Fig. 1). The open reading frames of each cDNA have a coding potential for 543 amino acids, seven amino acids less than that of the firefly luciferase cDNA. From a lysate of *E. coli* expressing click beetle luciferase, Western blot analysis revealed the presence of a single antigenic band which comigrates with the native enzyme (Wood *et al.*, 1988b). Since it is unlikely that a post-translational cleavage of the luciferase could occur at a common site in both beetles and bacteria, such cleavages probably do not occur in either host. However, the resolution of the blots is not sufficient to rule out deletions of less than about twenty amino acids. Other possible post-translational modifications to the luciferases are also unlikely to occur in either host for similar reasons. In the least, such modifications are limited to those that would not result in a significant alteration in the migration of the luciferases in SDS-polyacrylamide gel electrophoresis. The simplest explanation of the Western blot data is that the luciferases are expressed in either host directly as the mature enzymes from their mRNAs.

Analogous observations have been made previously in the comparison of native firefly luciferase with that produced from its cDNA in *E. coli*. In this case, it was also shown that firefly luciferase produced by *in vitro* translation of its

mRNA also comigrates with native enzyme (Wood *et al.*, 1984), this further indicating the absence of substantial post-translational modifications. However, some form of modification at the N-terminus of the native enzyme is implicated by our inability to obtain an amino acid sequence by Edman degradation. It is not at present known whether the N-terminus of firefly luciferase expressed in *E. coli* is also modified. Nevertheless, for both the firefly and click beetle luciferases, it is certain that there are no modifications required for enzymatic activity that are unique to the beetle metabolism, since the luciferases are active when expressed in exogenous hosts such as *E. coli*. Similarly, for the click beetle luciferases, variation in colour also is independent of possible beetle-specific modifications. Given the general lack of similarity between beetle and *E. coli* metabolisms, it is unlikely that there are any essential modifications of any sort. Thus, we work from the premise that the effective structure of the luciferases is simply the folded polypeptide chains which are coded by the cDNA clones.

Comparison of the amino acid sequences of the four luciferases shows a high degree of similarity. Pairwise comparisons of the sequences reveal from 94% to 99% identity (Fig. 2). This is in contrast to their comparison with the firefly luciferase sequence, with which they are less than 50% identical (Wood *et al.*, 1988b). In terms of amino acids, the sequences differ from each other by 5 to 32 residues. This information can be used to construct an evolutionary tree of the luciferases (Fig. 3) (Fitch and Margoliash, 1967), which shows that the luciferases have evolved in the order of the colour they emit. The earliest of the click beetle luciferases is the green-emitting luciferase, and the latest is the orange-emitting luciferase. The amino acid differences between the click beetle luciferases are the only differences between the *E. coli* capable of expressing the different colours of bioluminescence. Therefore, this set of amino acid differences must be responsible for the differences in the colour of bioluminescence.

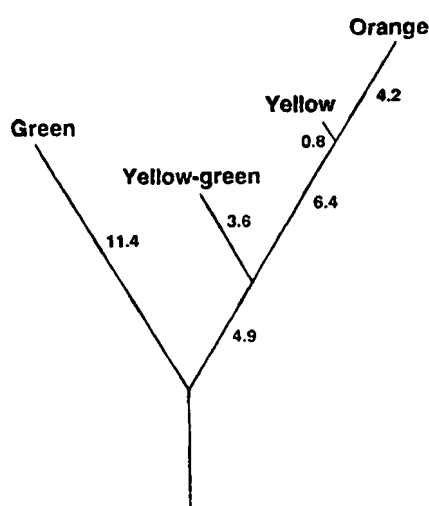
The effect that these amino acid differences have on the spectra of bioluminescence is very uniform. This can be demonstrated in a plot of the peak positions versus their respective widths at half maximal intensity for each of the four colours (Fig. 4). These parameters are presented in terms of wave numbers since this parameter is

Enzyme	Sequence	Position
BamHI	MMKREKNVIY GPEPLHPLED	50
Apal	LTAGEMLFRA LRKHSHPQA	100
BstXI	SYKEFFEATC LLAQSLHNCG	150
R	VAPVNESYIP DELCKVMGIS	200
A	LDTVENIHGC ESLPNFISRY	250
R V	LPGKVMQTHQ NICVRLIHAL	300
-V	GYFMVGLRVI MLRRFDQEAF	350
I	YDLSSLRELC CGAAPLAKEV	400
R- -	GDEFKSGSLG RVTPLMAAKI	450
-V T	NVEATKEAID DDGWLHSGDF	500
K	LEEILLKNPC IRDVAUVGIP	550
I	ERVSHTKYLR GGVRFVDSIP	600

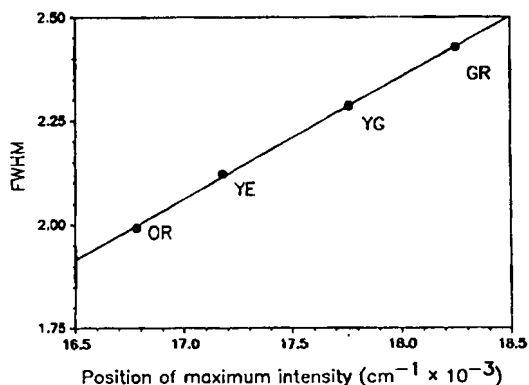
**Figure 1.** Amino acid sequences of the four click beetle luciferases capable of emitting four different colours. Shown in entirety is the sequence of the yellow-green-emitting luciferase. Above and below that sequence are shown only the amino acids at positions where the other luciferase sequences are different. Shown above the sequence are the amino acid differences for the green-emitting luciferase. Shown one line below the sequence are the amino acid differences for the yellow-emitting luciferase, and shown two lines below are the amino acid differences for the orange-emitting luciferase. A dash indicates no difference from the yellow-green emitting luciferase. Shown in bold type are the relative positions of restriction endonuclease site in the corresponding cDNA clones (the first BstXI site is not found in the green-emitting clone). The number on the right indicate the positions of the last amino acid in each line.

green	100%	96%	95%	94%
yellow-green	20	100%	97%	97%
yellow	23	11	100%	99%
orange	28	14	5	100%
	green	yellow-green	yellow	orange

**Figure 2.** Pairwise comparisons of the sequences of the four click beetle luciferases. The percentage identities of the sequences are shown in the upper right; the number of amino acid differences between the sequences are shown in the lower left (shaded area)



**Figure 3.** Evolutionary tree of the click beetle luciferases. Distances on the tree branches are given in number of amino acid differences between the sequences. The root of the tree was determined by comparison with the firefly luciferase



**Figure 4.** Plot of peak position versus peak width for each of the click beetle luciferases. Peak position is given as the wave number at which maximal intensity is found. Peak width is given as the width of the spectrum, in wavenumbers, found at half of the maximal intensity

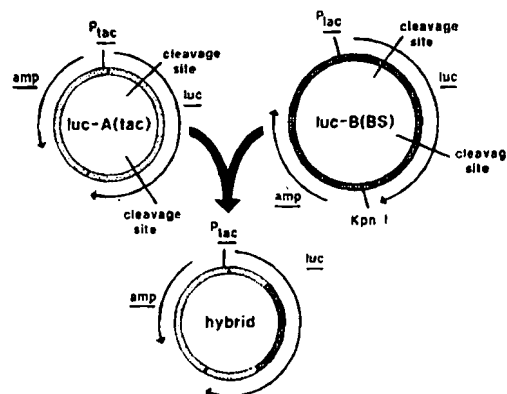
proportional to the energy of the photons emitted. In this form, the peak positions and widths are descriptive of the energy states of the light-emitting complex. The emitting complex refers to both the specific oxyluciferin fluorophore and its associations with the protein. The uniform manner in which these parameters vary suggests that discrete changes in the nature of the light-emitting complex are not effectuated to produce the alterations in colour. Hypothetically, such discrete changes could be tautomerization or ionization of the emitter, or coupling of the emitter to other conjugated  $\pi$ -bonded systems. Instead the predictable variation in the peak shapes is symptomatic of a modifying effect on the light emitter that can be varied in a continuous manner. Such modifying effects could be general changes in the dielectric constant around the emitter, or alterations of the electronic environment around specific key atoms in the emitter complex. The fact that the spectra of the four colours are nearly evenly spaced should not be construed to be a natural quantization of the system. Spectra taken from the ventral organ of live click beetles reveal other colours of bioluminescence, within the range of colours of the clones, which do not match any of the four colours from the clones (Biggley *et al.*, 1967). Presumably there are other members of this highly conserved group of click beetle luciferases that have not yet been cloned, which are capable of emitting different colours. Among all bioluminescent beetles there exists a large number of subtle shades in the colour of bioluminescence.

It should be remembered that the number of amino acid differences between the four sequences represents only an upper limit to the number required to produce the alteration in colour. As will be shown below, the number of amino acids that determine the colour of biolu-

minescence may be only a few. In the simplest *a priori* model, the colour would be determined by the chemical properties of four different amino acids found as a *single* position in all four luciferase sequences. Changes in the characteristics of the amino acids at this position, such as charge or hydrophobicity, would modify the colour of emission from one clone to the next in the uniform manner observed. However, with only the limited set of twenty amino acids available, it is difficult to find a uniform procession of chemical properties that could account for all four colours. The situation is worse when also included in the model are the other colours from the click beetle luciferases that have not yet been cloned. The fact is, comparison of the click beetle sequences reveals that there is no position in all four luciferases which changes to more than one other amino acid. Thus, as is typically the case, the role of individual amino acids to enzymatic activity is more subtle than the simplest model would predict.

#### METHOD FOR CONSTRUCTION OF HYBRID LUCIFERASES AND SPECTRAL ANALYSIS

Site-directed mutagenesis has become the standard method for determining the function of specific amino acids within an enzyme. However, to be practical, the method requires knowledge of which amino acids to modify. In our case, there are a total of 31 sites which are variant between the four luciferase sequences. To mutagenize each one would be very time-consuming and costly. As initial experiments, to reduce the number of options, we chose a method that would allow us to change several amino acids simultaneously (Fig. 5). Since the luciferase sequences vary by only a small amount, most of the restriction endonuclease cleavage sites in the corresponding cDNA sequences are conserved in all four clones. We are therefore able to express hybrid luciferases in *E. coli* by cleaving the cDNA clones with restriction endonucleases, and recombining the DNA fragments in new arrangements. Often, though we are exchanging large regions of sequence information, we are only changing the identities of a few amino acids. By analysing the pattern of amino acid substitutions for a large number of such rearrangement hybrids, we hoped to identify a subset of key residues that affect the colour of bioluminescence. For this type of



**Figure 5.** Method of constructing hybrid luciferases. Luc-A(tac) is the *tac*-containing vector (upper left). Luc-B(BS) is the Bluescript-based vector (upper right). Each vector contains a luciferase cDNA (*luc*) and a penicillin resistance gene (*amp*). The hybrid luciferase cDNA is constructed by cutting each parent vector at the 'cleavage site', and recombining the appropriate fragments

analysis we assume that the variant amino acid position function independent of each other, and look for experimental evidence to the contrary. For the most part, this assumption has been consistent with our observations.

To allow verification afterwards that the rearrangement hybrids we were constructing were indeed made properly, we used a strategy involving two different types of plasmids. One plasmid is a small expression vector which contains the *tac* promoter. At present it is the vector which yields the greatest expression of bioluminescence. The other plasmid, Bluescript, was derived from the cloning vector with which the luciferase clones were originally isolated. This plasmid can be distinguished from the *tac*-containing vector by the presence of a unique KpnI restriction site. The construction is performed by cleaving both plasmids, each containing a different luciferase cDNA clone, into two fragments with a chosen set of restriction endonucleases. One fragment contains a portion of the luciferase cDNA and the entire plasmid vector. The other fragment is the remaining portion of the cDNA. A *tac* vector containing a hybrid cDNA is formed by ligating the vector-containing fragment from the parent *tac* vector, with its complementary fragment from the Bluescript vector. The incorporation into the final product of the correct vector-containing fragment, with its associated portion of the luciferase cDNA, can be quickly confirmed by the absence



of the KpnI site. The other fragment, which contains only luciferase sequence, must be confirmed by DNA sequencing since there are no unique restriction sites to identify one luciferase cDNA from another. The one exception to this is the green-emitting clone, which is lacking a BstXI site which is found in the other three cDNA clones.

The spectra of bioluminescence from *E. coli* containing click beetle luciferases can be measured either from intact cells or from cell lysates. It has been shown for each of the luciferases that the spectra measured by either method are identical (Wood *et al.*, 1988b). Since it is technically simpler and less time-consuming, we chose to measure our spectra directly from living cells. The cells are grown on nitrocellulose filters on top of nutrient agar. To initiate bioluminescence, the filters are removed from the agar and soaked with luciferin for 5 to 10 minutes as described (Wood and DeLuca, 1987). The filters are then blotted dry and placed in the spectrometer for measurement.

The spectrometer is a Fastie-Ebert type as described (Seliger *et al.*, 1964). The output of the photomultiplier tube is amplified and digitally converted for direct input into an IBM PC compatible computer. Each sample was scanned five times at a rate of approximately 8 nm/s over a total period of 5 minutes. Noise in the spectra was reduced by applying a digital curve-smoothing routine to the data. The set of five spectra was then corrected for the wavelength-dependence sensitivity of the photomultiplier tube, and for time-dependent variation in the intensity of the bioluminescent sample. The final spectra was then computed as the average of the five original spectra. The precision of this method varies, dependent on intensity of the sample. Except for very weak samples, there is a greater than 95% confidence that two spectra are different if their maxima differ by 2 nm or more. This confidence is greater for the more intense samples. For very weak samples, several sets of spectra may be averaged together to reduce the effects of noise.

#### IDENTIFICATION OF SPECIFIC AMINO ACIDS WHICH AFFECT THE COLOUR OF BIOLUMINESCENCE

By using different combinations of restriction enzymes on the cDNA clones, the four sequences

were recombined to form four sets of rearrangement hybrids with five to eight members per set (Fig. 6). These hybrids, combined with the four original clones, form a total of 31 different luciferase sequences. Fig. 1 shows the relative positions of restriction sites which were used in the amino acid sequences of the luciferases. Also used in the constructions were two sites in the vectors near the ends of the cDNA inserts: BamHI at the 5'-end, and XhoI at the 3'-end. The first set of hybrids was constructed using NcoI and BamHI. This separates the last two variant amino acid positions from the remainder of the sequences. The second set of hybrids was constructed using ApaI and XhoI, which separates the first two variant positions from the remainder. The two BstXI sites, which separate approximately the central third of the variant positions, were used to construct the third set. This could not be done with the green-emitting clone because the first BstXI site was not present. This fact was used in the construction of the fourth set, which used the BstXI sites and the BamHI site to rearrange the final third of the variant position. For those clones which contained two BstXI sites, this resulted in cleavage of the fragment which contains only luciferase sequence into two. These fragments were combined and treated in the procedure as if they were only one, which was possible because the BstXI sites are not self-compatible. Thus, in the final ligation, the fragments can only recombine in one orientation.

To simplify nomenclature, we refer to the green-emitting clone as luc-GR, the yellow-green-emitting clone as luc-YG, the yellow-emitting clone as luc-YE, and the orange-emitting clone as luc-OR. The hybrids are referred to as 'luc-' followed by a number and a letter. The number refers to the set of hybrids which the clone is derived from, and the letter is an arbitrary distinction between the members of the set. Thus luc-2d is a hybrid that came from the second set of hybrids. A group of amino acid substitutions are referred to as a set of changes, and are indicated as shown in this example: [R<sub>223</sub>, L<sub>238</sub> → E, V] indicates that arginine at position 223 changes to glutamate, and that leucine at position 238 changes to valine. The inverse set would be changes in the opposite direction, i.e. [G<sub>223</sub>, V<sub>238</sub> → R, L]. The amino acid substitutions required to change one luciferase into another are indicated as shown: luc-GR → 2b [V<sub>9</sub>, L<sub>21</sub> → I, K] indicates the necessary amino acid substitutions

**Wild Type**

luc-GR	GR	VL.LYEWIT.SVGRLDARVRVL.SDVIIRDSI.EI	546 nm
luc-YG	YG	IL.LFDSL.A.NISKINTQIRAL.SDVIVGDSV.EK	560 nm
luc-YE	YE	IK.IFDSL.A.NVSKINTQIEAV.SEIVIGDSI.EI	578 nm
luc-OR	OR	IK.IFDSL.A.NISKINTQIEAV.GEIVIGVTI.KI	593 nm

**x NcoI, BamHI hybrids**

luc-1a	GR*YG	VL.LYEWIT.SVGRLDARVRVL.SDVIIRDSI*EK	546 nm
luc-1b	GR*OR	VL.LYEWIT.SVGRLDARVRVL.SDVIIRDSI*KI	546 nm
luc-1c	YG*GR	IL.LFDSL.A.NISKINTQIRAL.SDVIVGDSV*EI	559 nm
luc-1d	YG*OR	IL.LFDSL.A.NISKINTQIRAL.SDVIVGDSV*KI	561 nm
luc-1e	YE*YG	IK.IFDSL.A.NVSKINTQIEAV.SEIVIGDSI*EK	579 nm
luc-1f	YE*OR	IK.IFDSL.A.NVSKINTQIEAV.SEIVIGDSI*KI	578 nm
luc-1g	OR*GR	IK.IFDSL.A.NISKINTQIEAV.GEIVIGVTI*EI	593 nm
luc-1h	OR*YG	IK.IFDSL.A.NISKINTQIEAV.GEIVIGVTI*EK	593 nm

**x ApaI, XhoI hybrids**

luc-2a	YG*GR	IL*LYEWIT.SVGRLDARVRVL.SDVIIRDSI.EI	546 nm
luc-2b	OR*GR	IK*LYEWIT.SVGRLDARVRVL.SDVIIRDSI.EI	546 nm
luc-2c	GR*YG	VL*LFDSL.A.NISKINTQIRAL.SDVIVGDSV.EK	561 nm
luc-2d	OR*YG	IK*LFDSL.A.NISKINTQIRAL.SDVIVGDSV.EK	560 nm
luc-2e	GR*YE	VL*IFDSL.A.NVSKINTQIEAV.SEIVIGDSI.EI	579 nm*
luc-2f	YG*YE	IL*IFDSL.A.NVSKINTQIEAV.SEIVIGDSI.EI	578 nm
luc-2g	GR*OR	VL*IFDSL.A.NISKINTQIEAV.GEIVIGVTI.KI	593 nm*
luc-2h	YG*OR	IL*IFDSL.A.NISKINTQIEAV.GEIVIGVTI.KI	593 nm

**x BstXI hybrids**

luc-3c	OR*YG*OR	IK.IFDSL.A*NISKINTQIRAL*GEIVIGVTI.KI	580 nm
luc-3d	YE*YG*YE	IK.IFDSL.A*NISKINTQIRAL*SEIVIGDSI.EI	563 nm
luc-3e	YG*OR*YG	IL.LFDSL.A*NISKINTQIEAV*SDVIVGDSV.EK	577 nm
luc-3f	YE*OR*YE	IK.IFDSL.A*NISKINTQIEAV*SEIVIGDSI.EI	578 nm
luc-3g	YG*YE*YG	IL.LFDSL.A*NVSKINTQIEAV*SDVIVGDSV.EK	577 nm
luc-3h	OR*YE*OR	IK.IFDSL.A*NVSKINTQIEAV*GEIVIGVTI.KI	594 nm

**x BstXI, BamHI hybrids**

luc-4a	GR*YG	VL.LYEWIT.SVGRLDARVRVL*SDVIVGDSV.EK	550 nm
luc-4b	GR*YE	VL.LYEWIT.SVGRLDARVRVL*SEIVIGDSI.EI	553 nm
luc-4c	GR*OR	VL.LYEWIT.SVGRLDARVRVL*GEIVIGVTI.KI	571 nm
luc-4d	YG*GR	IL.LFDSL.A.NISKINTQIRAL*SDVIIRDSI.EI	559 nm
luc-4e	OR*GR	IK.IFDSL.A.NISKINTQIEAV*SDVIIRDSI.EI	573 nm

**Figure 6.** Sequence of hybrid luciferases. Each set of hybrids is indicated in bold type by the restriction endonuclease sites used in its construction. Each line thereafter corresponds to a different hybrid within the set. The first entry of each line is the hybrid name. The second entry indicates the origin of the fragments used to construct the hybrid shown in their correct order. The third entry shows, in succession, the amino acids at every position in the four luciferase sequences where there is at least one difference between the sequences. The '\*' in the series of amino acids indicate the positions of the restriction endonuclease sites. The restriction sites of actual use for the construction of a particular set of hybrids are indicated by '\*'. The last entry of each line indicates the wavelength at which the intensity maxima occurs for the spectra of each hybrid luciferase. Maxima values marked with an asterisk are from spectra that were very weak, so that the exact value is somewhat uncertain.

required to change the green-emitting luciferase to hybrid 2b.

The results of the hybrid experiments can be best understood in terms of deviations from the native enzyme sequences for each luciferase. We will begin by examining luc-YG, whose spectra is most like that of the firefly luciferase. Changes to both the first two and the last two variant position did not produce any change in the spectrum of bioluminescence. These changes are luc-YG→2c [ $I_9 \rightarrow V$ ], luc-YG→2d [ $L_{21} \rightarrow K$ ], luc-YG→1c [ $K_{484} \rightarrow I$ ], and luc-YG→1d [ $E_{403}, K_{484} \rightarrow K, I$ ]. All

of these changes produced luciferases whose spectral maxima were at 560 nm (within the limits experimental error). By combining these sets, we can form a new set, identified as luc-YG→YG', which describes changes in luc-YG that have little or no effect on its spectrum. This new set is luc-YG→YG' [ $I_9, L_{21}, E_{403}, K_{484} \rightarrow V, K, K, I$ ]. For three members of this set, the substituting amino acids have very different chemical properties. The change of  $E_{403}$  to  $K_{403}$  is one of the only two substitutions possible between any of the luciferases where the charge on the amino acid

reverses. In luc-YG $\rightarrow$ 3e [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V] the spectrum shifts from 560 nm to 577 nm, which is indistinguishable from the spectrum of luc-YE. Luc-YG $\rightarrow$ 3g [I<sub>89</sub>,R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ V,E,V] has the same effect. The difference between these sets, [I<sub>89</sub> $\rightarrow$ V], therefore apparently has no effect on the spectrum and can be added to the set of substitutions which do not affect luc-YG, i.e. luc-YG $\rightarrow$ YG' [I<sub>9</sub>,L<sub>21</sub>,I<sub>89</sub>,E<sub>403</sub>,K<sub>484</sub> $\rightarrow$ V,K,V,K,I]. The difference between luc-YG $\rightarrow$ YE and luc-YG $\rightarrow$ 3e is [L<sub>21</sub>,L<sub>41</sub>,I<sub>89</sub>,D<sub>226</sub>,V<sub>282</sub>I<sub>283</sub>,V<sub>323</sub>,V<sub>389</sub>,K<sub>484</sub> $\rightarrow$ K,I,V,E,IV,I,I,I]. By subtracting luc-YG $\rightarrow$ YG' from this, the set [L<sub>41</sub>,D<sub>226</sub>,V<sub>282</sub>I<sub>283</sub>,V<sub>323</sub>,V<sub>389</sub> $\rightarrow$ I,E,IV,I,I] is formed. It may be inferred indirectly that this set also has little effect on the spectra of luc-YG since the spectra of luc-YE and luc-3e are virtually identical. This can be tested directly by luc-YG $\rightarrow$ 3d, which after removing the members of luc-YG $\rightarrow$ YG' forms this set inferred to have little effect on the spectrum. The spectral maxima of hybrid luc-3d is 563 nm, 3 nm higher than luc-YG. Thus, while this verifies that this set has little effect on the spectrum, it does have a measurable effect. This also demonstrates that it is both a necessary and sufficient condition that the amino acid substitutions which produce most of the spectral shift between luc-YG and luc-YE are in the set [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V].

As was found for the luc-YG, changes to the first two and last two variant positions of luc-YE have no effect on its spectrum. Combining luc-YE $\rightarrow$ 1e, luc-YE $\rightarrow$ 1f, luc-YE $\rightarrow$ 2e, and luc-YE $\rightarrow$ 2f, we get the set luc-YE $\rightarrow$ YE' [I<sub>9</sub>,K<sub>21</sub>,E<sub>403</sub>,I<sub>484</sub> $\rightarrow$ V,L,K,K]. Luc-YE $\rightarrow$ 3f [V<sub>89</sub> $\rightarrow$ I] also produces no shift in the spectrum and so, as with luc-YG, it can be included in luc-YE $\rightarrow$ YE' to form [I<sub>9</sub>,K<sub>21</sub>,V<sub>89</sub>,E<sub>403</sub>,I<sub>484</sub> $\rightarrow$ V,L,I,K,K]. Since luc-YE $\rightarrow$ OR [V<sub>89</sub>,S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub>,E<sub>403</sub> $\rightarrow$ I,G,V,T,K] with the members of luc-YE $\rightarrow$ YE' removed is [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T], the amino acid substitutions that shift the spectra of luc-YE to luc-OR must be in this subset. By reasoning similar to that used to identify the set which can shift the spectra of luc-YG to luc-YE, the inverse of this set must contain the amino acid substitutions which can shift the spectra of luc-YE to nearly to luc-YG, i.e. [E<sub>223</sub>,V<sub>238</sub> $\rightarrow$ R,L] shifts the spectra of luc-YE to 563 nm.

The spectral shifts caused by [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T] and [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V] can be shown to act largely independently of each other by examining luc-YG $\rightarrow$ OR [L<sub>21</sub>,L<sub>41</sub>,

R<sub>223</sub>,L<sub>238</sub>,S<sub>247</sub>,D<sub>266</sub>,V<sub>282</sub>I<sub>283</sub>,V<sub>323</sub>,D<sub>352</sub>,S<sub>358</sub>,V<sub>389</sub>,E<sub>403</sub>,K<sub>484</sub> $\rightarrow$ K,I,E,V,G,E,IV,I,V,T,I,K,I]. After removing the members of luc-YG $\rightarrow$ YG', this set can be divided into three subsets: [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V], [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T], and [L<sub>41</sub>,D<sub>266</sub>,V<sub>282</sub>I<sub>283</sub>,V<sub>323</sub>,V<sub>389</sub> $\rightarrow$ I,E,IV,I,I]. The first set is the set of substitutions shown to be largely responsible for the spectral shift of luc-YG to luc-YE, and the second set are those responsible for the spectral shift of luc-YE to luc-OR. The third set is the set shown to have a small but measurable effects on the spectra of luc-YG. Therefore, the spectral shift from luc-YG to luc-OR, 33 nm, appears to rely on both of the first two sets of effective substitutions. It can be shown that these two sets act independently by applying each to luc-YG. The effect of [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V] is already described by luc-YG $\rightarrow$ 3e and produces a spectral shift of 17 nm. Luc-YG $\rightarrow$ 3c, after removing members of luc-YG $\rightarrow$ YG' and members of the set which has a small effect on the spectra of luc-YG, is the set [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T]. Hybrid 3c is shifted 20 nm from luc-YG. The sum of the effects of [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V] and [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T] is 37 nm, about 10% greater than the spectral shift of luc-YG $\rightarrow$ OR. This is fairly good agreement, especially when acknowledging that luc-YG $\rightarrow$ OR has substitutions which contribute small effects that are not in either [R<sub>223</sub>,L<sub>238</sub> $\rightarrow$ E,V] or [S<sub>247</sub>,D<sub>352</sub>,S<sub>358</sub> $\rightarrow$ G,V,T]. Therefore, the effects of these two sets of mutations are largely additive.

Determining which amino acids are responsible for the green emission of luc-GR is a more difficult problem because the set of changes which describes this luciferase is so large. For example luc-YG $\rightarrow$ GR has 20 amino acid changes. The problem is compounded by the lack of one of the BstXI sites which have been so useful in analysing the other three luciferases. Hence, our analysis of luc-GR is so far only preliminary. The first two and last two variant positions have been found to have no effect on the spectrum of luc-GR, thus combining luc-GR $\rightarrow$ 1a, luc-GR $\rightarrow$ 1b, luc-GR $\rightarrow$ 2a, and luc-GR $\rightarrow$ 2b, gives luc-GR $\rightarrow$ GR' [V<sub>9</sub>,L<sub>21</sub>,E<sub>403</sub>,I<sub>484</sub> $\rightarrow$ I,K,K,K]. Luc-GR $\rightarrow$ 4a, with luc-GR $\rightarrow$ GR' removed, gives [I<sub>323</sub>,R<sub>351</sub>,I<sub>389</sub> $\rightarrow$ V,G,V], which shifts the spectrum of luc-GR 4 nm to 550 nm. Luc-GR $\rightarrow$ 4b, with members of luc-GR $\rightarrow$ GR' removed, gives [D<sub>226</sub>,V<sub>282</sub>I<sub>283</sub>,R<sub>531</sub> $\rightarrow$ E,IV,G] which shifts the spectrum to 553 nm. The difference between these two sets, i.e.

luc-4a→4b [ $D_{226}, V_{282}I_{283}, V_{323}, V_{389} \rightarrow E, IV, I, I$ ], shifts the spectrum 3 nm. This is a subset of a set identified earlier to have a 3 nm spectral shift in luc-YG→3d. Luc-GR→4c, after removal of luc-GR→GR', gives [ $S_{247}, D_{266}, V_{282}I_{283}, R_{351}, D_{352}, S_{358} \rightarrow G, E, IV, G, V, T$ ] which shifts the spectrum 25 nm to 571 nm. This set can be divided into [ $D_{266}, V_{282}I_{283}, R_{351} \rightarrow E, IV, G$ ] and [ $S_{247}, D_{352}, S_{358} \rightarrow G, V, T$ ]. The first set is the same as the effective set of luc-GR→4b, which gives a 7 nm shift, and the second set is the effective set of luc-YE→OR, which gives a 16 nm shift. Thus, the sum of these effects, 23 nm, is nearly the same as their combined affect. These observations are consistent with the interpretation that [ $I_{323}, R_{351}, I_{389} \rightarrow V, G, V$ ] can effect a 4 nm shift in the spectrum independent of other amino acid changes. As an additional example, luc-4e→3g, with removal of the amino acid changes shown to not affect the spectra, gives [ $L_{41}, I_{323}, R_{351}, I_{389} \rightarrow I, V, G, V$ ], which causes a spectral shift 4 nm, from 577 nm to 573 nm. However, [ $I_{323}, R_{351}, I_{389} \rightarrow V, G, V$ ] can also be formed from luc-4d→YG with members of luc-YG→YG' removed. But in this case there is almost no change in the spectrum.

## SUMMARY AND CONCLUSIONS

We have generated four different types of cDNA clones from the ventral light organ of *Pyrophorus plagiophthalmus*. These clones can direct the synthesis of luciferase in *E. coli* which are distinguishable by the colour of bioluminescence they emit. Since the different colours are expressible in a bacterial host, they cannot be due to post-translational modifications that are unique to beetles. Further, because of the many differences between prokaryotes and eukaryotes, it is unlikely that there are any post-translational modifications responsible for the different colours. Thus the determinants of colour must be found among the amino acid differences in the sequences of the four luciferases. We have begun to identify these determinants by constructing and analysing hybrid luciferases, made by recombining fragments of the four different types. The results have shown that there are two groups of amino acids that each can produce a greater than

16 nm change in the spectrum of a luciferase. There are at least two other groups of amino acids that can cause smaller changes in the spectra, and several amino acid which have virtually no effect on the spectra. The two sets of amino acid can be expressed as directional substitutions, [ $R_{223}, L_{238} \rightarrow E, V$ ] and [ $S_{247}, D_{352}, S_{358} \rightarrow G, V, T$ ], which in this form result in a large spectral shift towards longer wavelengths. It was shown that the effect of these two set of changes can act largely independent of each other. For the other sets of changes that cause smaller effects, the independence of their action is less clear.

It is not known by this analysis whether all amino acids in each set are required to effectuate a change. To determine this we are initiating site-directed mutagenesis to produce changes at only single sites. We anticipate that the amino acids producing the largest effect on the spectra will be those which have the largest changes in their chemical properties. Thus, in the set [ $R_{223}, L_{238} \rightarrow E, V$ ], the effect is probably attributable mostly to  $R_{223} \rightarrow E$ , which is a reversal of charge. In the set, [ $S_{247}, D_{352}, S_{358} \rightarrow G, V, T$ ], the effect is probably due to  $D_{352} \rightarrow V$ , which changes a negative charge to a hydrophobic residue. But  $S_{247} \rightarrow G$  may also be effective since it is a loss of hydrogen bonding.

## REFERENCES

- Biggley, W. H., Lloyd, J. E. and Seliger, H. H. (1967). The spectral distribution of firefly light II. *J. Gen. Physiol.*, **50**, 1681-1692.
- Fitch, W. M. and Margoliash, E. (1967). Construction of phylogenetic trees. *Science*, **15**, 279-284.
- Seliger, H. H., Buck, J. B., Fastie, W. G. and McElroy, W. D. (1964). The spectral distribution of firefly light. *J. Gen. Physiol.*, **48**, 95-104.
- Wood, K. V. and DeLuca, M. (1987). Photographic detection of luminescence in *Escherichia coli* containing the gene for firefly luciferase. *Anal. Biochem.*, **161**, 501-507.
- Wood, K. V., de Wet, J. R., Dewji, N. and DeLuca, M. (1984). Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns. *Biochem. Biophys. Res. Commun.*, **124**, 592-596.
- Wood, K. V., Lam, Y. A., Seliger, H. H. and McElroy, W. D. (1988a). cDNA's encoding click beetle luciferases can elicit bioluminescence of different colors. (Submitted for publication).
- Wood, K. V., Lam, Y. A. and McElroy, W. D. (1988b). Introduction to beetle luciferases and their applications. *J. Biolum. Chemilum.* (in press).

## Introduction to Beetle Luciferases and their Applications

Keith V. Wood\*, Y. Amy Lam and William D. McElroy

Department of Chemistry, M-001, University of California, San Diego, La Jolla, CA 92093, USA

All beetle luciferases have evolved from a common ancestor: they all use ATP, O<sub>2</sub>, and a common luciferin as substrates. The most studied of these luciferases is that derived from the firefly *Photinus pyralis*, a beetle in the superfamily of Cantharoidea. The sensitivity with which the activity of this enzyme can be assayed has made it useful in the measurement of minute concentrations of ATP. With the cloning of the cDNA coding this luciferase, it has also found wide application in molecular biology as a reporter gene. We have recently cloned other cDNAs that code for luciferases from the bioluminescent click beetle, *Pyrophorus plagiophthalmus*, in the superfamily Elateroidea. These newly acquired luciferases are of at least four different types, distinguishable by their ability to emit different colours of bioluminescence ranging from green to orange. Unique properties of these luciferases, especially their emission of multiple colours, may make them additionally useful in applications.

**Keywords:** Firefly luciferase; click beetle luciferases; reporter genes

### INTRODUCTION

Man's perception of the world is visually oriented. Since bioluminescence is one of the few things that can be seen in the dark, it is understandable that this has been a topic of biochemical research for many decades. Fireflies have been prominent in this research endeavour, in part because they are abundant and their light organs are replete with luciferase. Thus they provided a plentiful resource for further study. Early research on fireflies was done primarily to better understand this peculiar phenomena of living light. Sometimes, though, the earliest work was justified as a means of developing artificial lighting. In the late 1940s, when the general

importance of ATP metabolism was just becoming recognized, it was discovered that ATP was a component in the luminescent reaction of fireflies. The firefly luciferase became one of the paradigms of ATP-utilizing enzymes. Because of the extreme sensitivity with which the activity of this enzyme could be assayed, it was soon adapted as a tool in the measurement of very low concentrations of ATP. Subsequently, luciferase was combined with other ATP-utilizing enzymes to produce coupled enzymatic systems. In these systems, the luciferase was the reporter allowing sensitive measurements of a wide variety of metabolites.

Recently, firefly luciferase has found new application as a reporter of genetic activity in

\*Author for correspondence.

living cells. Its application in this area was made possible by the cloning of its cDNA, which can direct the synthesis of active enzyme in foreign hosts (de Wet *et al.*, 1985). In this report, we describe briefly how the properties of luciferase have made it well suited for this purpose. We also present information on the recent cloning of several new cDNAs from a bioluminescent click beetle (Wood *et al.*, 1988a). These cDNA clones encode for four different types of luciferases, which can be distinguished by their ability to emit different colours of bioluminescence. These colours range from green to orange. Structurally, the click beetle luciferases differ significantly from the firefly luciferase, and these differences are reflected in their chemical properties. Because of this, the click beetle luciferases may have additional features to make them useful as genetic reporter.

#### PROPERTIES OF THE FIREFLY LUCIFERASE

All beetle luciferases catalyse the conversion of chemical energy into light by a two-step process (Fig. 1) (Seliger and McElroy, 1964; DeLuca and McElroy, 1978). This process utilizes ATP,  $O_2$ , and beetle luciferin, a unique heterocyclic acid found only in bioluminescent beetles. In the first step, the carboxylate group of luciferin is activated by acylation with the alpha-phosphate of ATP. The luciferyl adenylate is then oxidized with molecular oxygen, in the second step, to yield AMP, carbon dioxide, and oxyluciferin. The oxyluciferin is generated in an electronically excited state which, upon transition to the ground state, emits the photon characteristic of bioluminescence. For firefly luciferase, the most studied of the beetle luciferases, the quantum yield for this reaction has been measured at 0.88 relative to the consumption of luciferin (McElroy

and Seliger, 1960). This is the highest yield reported for any luminescent reaction.

Under optimal conditions the firefly luciferase emits light whose peak intensity is at 561 nm (yellow-green). This is the same as the light emitted from live fireflies. Under a variety of conditions, however, the structure of luciferase can be altered to a form which emits predominantly at 617 nm (red) (Seliger and McElroy, 1964). Some conditions which can cause this spectral shift are pH below 7.5, temperature above 20 °C, the presence of denaturants such as urea, and the presence of heavy metals such as  $Zn^{2+}$ ,  $Cd^{2+}$ , or  $Hg^{2+}$  (Seliger and McElroy, 1964). Some chemical modifications of the enzyme, or the use of substrate analogues, can also cause the enzyme to emit red light (DeLuca *et al.*, 1973; Alter and DeLuca, 1986). In the case of pH, the shift to red light is associated with a substantial decrease in the quantum yield of the reaction (McElroy and Seliger, 1966). This decrease in quantum yield is probably evident under any condition that promotes the red-emitting form. The spectral shift associated with changes in temperature, or the presence of denaturants, can be interpreted as resulting from partial unfolding of the enzyme structure. For others conditions, it is not known whether the effects are localized to key reactive residues, or whether they also cause general perturbations to the structure. Aside from the actual decrease in the quantum efficiency of luciferase in the red-emitting form, the spectral shift also causes an apparent decrease in enzymatic activity. This is because photomultiplier tubes are generally much less sensitive to red light than green light. Both these real and apparent effects combine to give a large pH dependence to the measured light output of firefly luciferase. The optimal pH for light output is pH 7.8.

Under conditions of excess substrates, the light output of luciferase is proportional to the

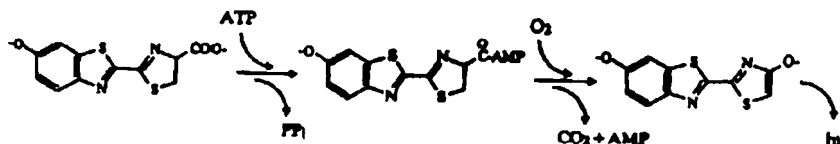


Figure 1. The luminescent reaction catalysed by beetle luciferase. Beetle luciferin is the heterocyclic molecule depicted on the left. Other reactants are depicted by their common abbreviations. The final product,  $hv$ , is the photon of light emitted from oxyluciferin during its transition to its ground-state electronic configuration.

concentration of enzyme over at least a 10,000-fold range. With the use of a sensitive luminometer, as little as 10 femtograms of enzyme can be detected ( $10^7$  molecules). Initiation of the luminescent reaction by rapid mixing of the substrates with the enzyme results in a rapid release of light which reaches maximum intensity after about 0.3 seconds. The intensity is then quickly inhibited in a biphasic manner, reaching about 10% of its peak value after 30 seconds. Beyond one minute the intensity is produced from a steady-state process which decays slowly over several hours. Because the peak luminescence is typically over 10-fold greater than the steady-state luminescence, the enzymatic assay is most sensitive when the first 30 seconds of the reaction are included in the measurement of total light output. This is typically accomplished in a luminometer where the reaction is initiated in front of the photomultiplier tube by injection of substrates directly into the sample holder.

Firefly luciferase is a 61 kD enzyme which apparently is active as a monomer. It is coded by a 550 amino acid reading frame in its mRNA, and is probably produced as an active enzyme without the necessity of post-translational modifications. In the firefly, the enzyme is located in specialized peroxisomes of the light organ (Keller *et al.*, 1987). It is directed to these sub-cellular organelles by a targeting sequence at the C-terminus of the protein (Gould and Subramani, 1988b, 1988c). This targeting sequence is conserved throughout eukaryotes, and will cause the luciferase to localize in the peroxisomes of other organisms when expressed in exogenous hosts.

#### APPLICATION OF FIREFLY LUCIFERASE AS A REPORTER GENE

With the past decade have come dramatic advancements in our ability to manipulate genetic materials. Enabled by these new techniques, the study of sub-cellular events which regulate genetic activity has become one of the largest areas of research today. A key tool in this area of research is the reporter gene, which provides an observable parameter in the monitoring of genetic events at a molecular level. In its simplest form, a reporter gene is a fragment of DNA which encodes for an easily detectable protein. This protein is the reporter. In experiments, the reporter gene is linked to other fragments of

DNA which are thought to contain genetic control elements, and the assemblage is introduced into living cells. Production of the reporter in the cell is regulated by the action of the control elements on the transcriptional activity of the reporter gene. Thus, the reporter is the observable parameter allowing the experimenter to monitor the action of the control elements.

In practice, the transcriptional activity of a reporter gene can be quite low, and experiments are often limited by an inability to detect the reporter. Therefore, for a reporter to be widely useful, it must be detectable in very low concentrations. In addition, the reporter must be detectable by a method that can distinguish it from other proteins native to the host cell. Firefly luciferase meets these criteria ideally. It can be detected in very small amounts through its bioluminescent activity, and since bioluminescence is not a common event in living systems, its activity will be unique in the experimental host. That is, there is no endogenous luminescent activity of the host to interfere with the detection of even miniscule amounts of luciferase. The bacterial enzyme, chloramphenicol acetyltransferase (CAT), has been used conventionally as a reporter in eukaryotic systems for similar reasons. Its enzymatic activity is not found in eukaryotic cells, so CAT can also be detected without confusion from host activities. Its assay is based on conversion of the antibiotic chloramphenicol to mono- and di-acetylated forms. High sensitivity is provided by the use of  $^{14}\text{C}$ -labelled chloramphenicol as the substrate. This method requires that the products of the reaction be separated from the substrate before quantification, usually by thin layer chromatography or HPLC.

Because CAT is widely used as a genetic reporter, it was used as a benchmark to evaluate the suitability of firefly luciferase in this application (de Wet *et al.*, 1987). It was found that the levels of expression of CAT and luciferase in eukaryotic systems were comparable. It had been previously shown that the production of CAT in eukaryotic cells is proportional to mRNA transcription from the reporter gene. Since luciferase production paralleled CAT production under a variety of experimental conditions, luciferase must also be a proportional indicator of transcriptional activity. However, because of the efficient detection methods achievable with bioluminescence, the assay of luciferase is 100 to 1000 times

more sensitive. Thus, much lower levels of genetic activity are detectable. Furthermore, the time required to assay luciferase is much less than CAT. Using a luminometer or scintillation counter, the luciferase assay requires about a minute per sample. The CAT assay usually requires several hours. The assay of luciferase also does not require the special precautions needed for radioactive  $^{14}\text{C}$ .

One of the unique advantages of firefly luciferase as a reporter of genetic activity is its potential to measure this activity from within living cells. This is not possible with use of CAT since the products of the reaction require separation from the assay mixture in order to be quantified. The photons produced in the luciferase reaction, however, are generally able to pass from within the host cell to allow external detection. A precondition of this is that the luciferin substrate be able to pass into the cell to combine with the luciferase reporter. The other substrates of the reaction, ATP and  $\text{O}_2$ , are readily available in the interior of the cell. The mere addition of luciferin to the external media is sufficient to allow its passage across the cellular membrane. But the light output elicited by this method is less than expected given the extent of luciferase contained within the cells. Light output can be increased with the use of permeabilizing agents such as DMSO or nigericin (Gould and Subramani, 1988a), but still not to the full potential expected. It is not known whether permeability of the outer membrane is the only limitation, or whether there are other inhibitors of activity. One possibility is that the peroxisomal membrane acts as a second barrier to luciferin passage. Since luciferase is localized into peroxisomes, most of the luminescent activity may arise from these organelles. Experiments are currently under way to remove the peroxisomal targeting signal from luciferase so that it will remain in the cytoplasm. This may improve its ability to elicit luminescence from within intact cells. However, other possibilities exist, such as unfavourable microenvironmental effects, which could inhibit the activity of luciferase in a foreign host.

Since the first published reports of its use as a genetic reporter, this new application of firefly luciferase has received much interest. By the time this article was written, we had received approximately 1000 requests for the cDNA encoding luciferase from other laboratories wishing to apply it to their experimental systems. The

feedback from these other laboratories has been quite positive. In most cases, researchers are finding that the use of luciferase instead of CAT is saving much time in the execution of their experiments. The time saved is not only in the much shorter assay time of luciferase, but also in the time required for sufficient expression of the reporter. Previously, production of the reporter often was not detectable for 24 to 48 hours after the reporter gene was introduced into cells. Because of the much higher sensitivity of the luciferase assay, expression of the reporter gene can typically be measured after only a few hours. In some cases, where expression of the reporter was previously too low for detection under any conditions, the use of luciferase has allowed measurements to be made. To date, luciferase has been expressed from its cDNA in almost every living kingdom. It has been expressed in bacteria (de Wet *et al.*, 1985), yeast (Wood and DeLuca, 1987), dictyostelium (Howard *et al.*, 1988), mammalian cells (de Wet *et al.*, 1987; Gould and Subramani, 1988a), and plant cells (Ow *et al.*, 1986), as well as in transgenic mice (DiLella *et al.*, 1988; Crenshaw and Rosenfeld, 1988) and plants (Ow *et al.*, 1986).

#### COMPARISON OF FIREFLY AND CLICK BEETLE LUCIFERASES

Bioluminescent beetles are found in two superfamilies, Elateroidea and Cantharoidea (Lloyd, 1978). Fireflies are members of the family Lampyridae in the superfamily Cantharoidea; as indicated above, they have been the primary source of a beetle luciferase because of their abundance and accessibility. In the superfamily Elateroidea, only the family Elateridae contains bioluminescent members, which are more commonly known as click beetles. This family of beetles is one of the most widely distributed, with species found in most areas of the world. However, unlike Lampyridae, where nearly all of the members are bioluminescent, only a small percentage of Elateridae are so. Most of these are located in the Caribbean and in South America. Their taxonomy suggests that the click beetles are the most evolutionarily distant of the bioluminescent beetles from the fireflies (Crowson, 1981). The time of divergence of the Elateroidea and Cantharoidea superfamilies cannot be estimated directly owing to a lack of fossils. But by



comparison of the morphological differences between these groups of beetles, corroborated with the fossil record of other beetles, it has been estimated that these superfamilies diverged about 120 million years ago.

Morphologically the click beetles and fireflies are quite distinct (Fig. 2). The click beetles have a hard exoskeleton, and are often larger than the fireflies. They can be recognized by a characteristic behaviour they display when being constrained or placed on their backs. They make an audible 'click' sound by forcibly arching their head forward. Bioluminescent click beetles have two sets of light organs. One pair is located on the dorsal surface of the head. These light organs emit long pulses of light when the beetles are not in flight. The second set is a single organ located in a cleft on the ventral surface of the beetle between the mesothorax and abdomen. This light organ also emits long pulses of light but only when the beetle is in flight. On the ground the cleft is closed and the light is extinguished. For most species of bioluminescent click beetle, the ventral organ emits light at a longer wavelength than the dorsal organ. The position and activity of the light organs in fireflies is quite different. These beetles have one set of light organs located on the ventral surface of the abdomen, on the posterior sternites. They gener-

ally emit short burst of light in a pattern which is indicative of the particular species.

One species of click beetle has been of particular interest since its bioluminescence was first studied in 1963. This species, *Pyrophorus plagiophthalmus*, is an especially large click-beetle being typically 3cm in length. It was of interest because its bioluminescence presents an unusually large range of colour (Seliger *et al.*, 1964). Furthermore, the colours vary between individuals, a property not found in fireflies. The light of the dorsal organ is greenish in colour, but varies between individuals from green (548 nm) to yellow-green (565 nm). The ventral organ varies over a much wider range, from green (547 nm) to orange (594 nm) (Biggley *et al.*, 1967). The luciferases of these beetles were extracted to determine the source of these different colours. In extracts, the bioluminescence spectra were not different from those of the living beetles. Analysis showed that the different colours were not due to alterations of the substrates of the reaction, which are the same as utilized by the firefly luciferase. It was concluded that the differences were due to variation in the interaction of the substrates with the luciferases (Seliger and McElroy, 1964). Unfortunately, attempts to study these luciferases further were limited by the difficulty of collecting sufficient quantities of the beetles.

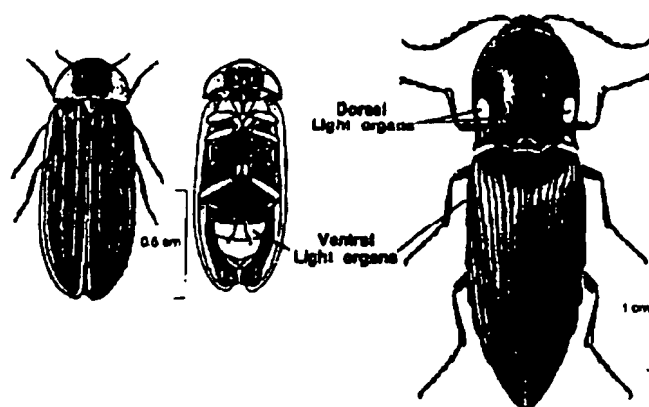


Figure 2. The general morphology of fireflies and click beetles. A firefly (Lampyridae) is shown in two views on the left, and a click beetle (Elateridae) is shown on the right.

### CLONING AND EXPRESSION OF CDNAS ENCODING CLICK BEETLE LUCIFERASES

Expression of firefly luciferase in *Escherichia coli* demonstrated that we could produce this enzyme from a source which was easily grown in the laboratory. Thus, in this case, the information contained in the cDNA encoding firefly luciferase was in itself sufficient to generate an active enzyme in a foreign host. Application of this technology to the click beetle luciferases could circumvent the problems of collecting large quantities of the beetles. The methods used in cloning a cDNA which encodes luciferase require only a small supply of the beetles, and they are needed only once. Afterwards, bacterial hosts generate the DNA and enzyme needed for further study. Production of the click beetle luciferases from cDNA clones has the additional advantage that genetic variants of the enzyme, such as those which produce the different colours of bioluminescence, are generated in isolation of one another. Enzymes isolated from the click beetles directly would require methods for separation of the different variants. This would be difficult since, as was subsequently found, the physical differences between these variants are few. Furthermore, the amino acid sequences of the luciferases can be determined from the DNA sequences of their cDNA clones. DNA sequencing is an efficient technique making it practicable to determine the amino acid sequence differences between several proteins of over 500 amino acids each.

Specimens of *Pyrophorus plagiophthalmus*, collected from the northeast end of Jamaica, were transported live to the laboratory and frozen in liquid N<sub>2</sub>. Messenger RNA was isolated from ventral light organs of approximately 60 beetles, one microgram of which was converted to cDNA (de Wet *et al.*, 1986). This was packaged in a specialized lambda cloning vector, Lambda ZAP, to yield 700,000 recombinant plaques (Fig. 3) (Short *et al.*, 1988). We had originally intended to screen the library by DNA hybridization using the cDNA sequence of firefly luciferase. However, attempts at visualizing the click beetle luciferase gene in Southern blots, using the firefly luciferase cDNA as the probe, failed to demonstrate cross-hybridization even under conditions of low stringency. It had been previously shown that antibodies raised against firefly luciferase can cross-react with the click beetle luciferases

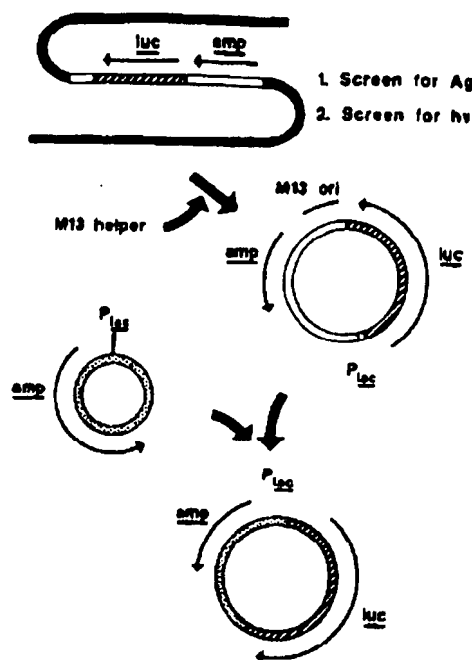


Figure 3. The strategy for cloning and expressing the cDNAs coding click beetle luciferases. The upper left corner depicts the Lambda ZAP vector. The cDNA library in this form was screened for the expression of antigenic polypeptides (Ag). With the use of M13 helper, Lambda ZAP was transformed either individually or *en bloc* into Bluescript plasmids (upper right). The cDNA library in this form was screened for the expression of luminescence (hv). The expression of luminescence was improved by transferring the cDNA clones to a vector containing the tac promoter (lower left and bottom).

(Wienhausen and DeLuca, 1985). Thus we used such antibodies, raised in rabbits, to screen the cDNA library. Because the cross-reactivity with click beetle luciferases is weak, we used antibodies that had been affinity purified by selection on a Sepharose column containing immobilized firefly luciferase (a gift from Dr Gilbert Keller; Keller *et al.*, 1987).

The original screening was done on unamplified aliquots of the lambda cDNA library. Determined by the colorimetric detection of alkaline phosphatase conjugated to anti-rabbit antibodies, 5.5% of the recombinant lambda phage expressed luciferase antigens. Eighteen clones were chosen for further analysis. A unique

feature of the Lambda ZAP cloning vector is that it can be transformed into a bacterial expression plasmid (Bluescript) by an *in vivo* process involving the addition of an M13 helper phage (Short *et al.*, 1988). The recombinant lambda containing the four longest cDNA clones were transformed into their plasmid form, and found to be able to express bioluminescence in *E. coli*. It could be visually observed from the expression of the four clones in *E. coli* that two produced orange light, one produced yellow light, and one produced yellow-green light.

In order to ascertain whether other colours of bioluminescence could also be found in the library, it was rescreened for other full-length cDNA clones. The rescreening was done by a different method designed to identify luminescent activity directly. Five aliquots of the original library were amplified, then transformed *en bloc* into expression plasmids. As in the case of eukaryotic cell expressing the firefly luciferase (see above), bioluminescence can be initiated in *E. coli* expressing luciferase by the addition of luciferin to the media (Wood and Deluca, 1987). In bacteria, the diffusion of luciferin through the membranes can be facilitated by reducing the pH of the media to 5. Presumably this masks negative charges on the molecule, making it more hydrophobic and permeable to a lipid bilayer. By adding luciferin to bacterial colonies containing clones of the cDNA library, colonies able to express a functional luciferase were identified directly by their ability to darken X-ray film. Several bioluminescent colonies were isolated from each aliquot of the library, seven of which were identified as arising from independent cDNA clones. From two of the aliquots, two colonies could be judged as resulting from independent clones based on widely different intensities. The independence of these clones was later confirmed by restriction mapping. From these clones, five emit yellow light, one emits orange light, and one emits a new colour, green.

Western blot analysis was performed to confirm that full-length click beetle luciferases were being expressed in the *E. coli*. Despite the fact that some of these clones were clearly visualized by anti-firefly luciferase antibody during the library screening, we were unable to detect the gene products in blots made directly with *E. coli* lysates. This is the result of both a low level of gene expression, and a weak cross-reactivity with the antibody. The expression of luminescence was

increased by transferring the cDNA clones to a plasmid vector which incorporated a *tac* promoter (Fig. 3). A lysate from *E. coli* expressing the green-emitting luciferase from the *tac* vector further required partial purification to be detectable in a blot. The blot revealed a single band, cross-reactive with firefly luciferase, which comigrates with the native click beetle luciferase (Fig. 4). DNA sequence analysis was later performed



Figure 4. Western blot showing the expression of click beetle luciferase in *E. coli*. Lane 1: partially purified extract of *E. coli* expressing the green-emitting luciferase. Lane 2: extract of click beetle light organ. Lane 3: purified firefly luciferase. Luciferases were detected with anti-firefly luciferase.

for one clone of each colour. This confirmed that each cDNA contained an open reading frame which could code for a protein whose N-terminus corresponded to the N-terminus of firefly luciferase. Thus, as has been achieved previously with the firefly luciferase, the click beetle luciferases can be produced in *E. coli* as full-length and enzymatically active enzymes.

### BIOLUMINESCENCE SPECTRA OF CLICK BEETLE LUCIFERASES

Spectrographic analysis was performed on the bioluminescence emitted from *E. coli* expressing the various cDNA clones. Bioluminescence was induced from whole cells by the same method used previously in the screening of bacterial colonies for luminescence. Cells producing luciferase from the *lac* vector yielded sufficient light intensity, upon addition of luciferin to the media, to allow spectral measurements. These measurements verified the visual observation that the eleven clones can be sorted into four groups based on the colour of light emitted. For each of the colours, the spectrum is a single peak qualitatively similar to the spectra of native click beetle luciferase (Seliger *et al.*, 1964). When the spectra of the four colours are superimposed, they show a remarkable pattern of four similarly shaped peaks that are nearly evenly spaced (Fig. 5). The wavelengths of maximum intensity are 546 nm for green, 560 nm for yellow-green, 578 nm for yellow, and 594 nm for orange.

Spectra were also measured from lysates of the *E. coli* after partial purification (Wood *et al.*,

1988a) (Fig. 6). Bioluminescence was elicited from the lysates by diluting them 100-fold into a reaction mixture ranging in pH from 6 to 10. For pHs 6.0, 7.0, and 8.0, the reaction mixture was buffered with 50 mmol/l MES/50 mmol/l MOPS/50 mmol/l Tricine. For pHs 8.0, 9.0, and 10.0 it was buffered with 50 mmol/l Tricine/50 mmol/l CHES. Also in the reaction mixture were 5 mmol/l  $MgSO_4$ /1 mmol/l EDTA/0.1 mmol/l luciferin/1.5 mmol/l ATP/1 mmol/l NaF/0.2 mg/ml BSA/10% glycerol. (NaF was found to simplify the kinetics of the decay of luminescence, which simplified the analysis of the spectral data. It does not affect the spectral distribution. It is not known whether it affects the activity of the click beetle luciferase directly, or whether it is due to an interaction with other components of the lysate. It has no effect on the purified firefly luciferase.) For the click beetle luciferases from each of the four colours, the spectra measured from whole cells matched that of the lysates at pH 6.0 and pH 7.0. Also, for each of the luciferases, the spectra shifted towards longer wavelengths at pH above 9.0. This shift was largest for the green-emitting luciferase, less for the yellow-green-emitting luciferase, and the least for the yellow-green and orange-emitting luciferase. At pH 8.0, this shift is virtually undetectable for the yellow-green and orange-emitting luciferases. For the green and yellow-emitting luciferases, the shift at pH 8.0 can be detected as a slight widening of the spectral peak, but the position of the maxima is unchanged.

This pH response of the click beetle luciferases is in contrast with that of the firefly luciferase. As stated above, the spectrum of firefly luciferase shifts to longer wavelengths at low pH (Fig. 7). In the pH range of 8.0 to 10.0, the enzyme emits its characteristic yellow-green light. The spectrum shifts towards longer wavelengths at pH 7.0, and at pH 6.0 is generated almost completely from a red-light emitting form of the enzyme. This shift is much larger than is seen with the click beetle luciferases. At pH 7.0, where the spectrum of firefly luciferase is a mixture of yellow-green and red-emitting forms of the enzyme, a difference is apparent between the enzyme purified from fireflies and that produced in *E. coli* (Fig. 7). With the luciferase from *E. coli*, the red component of the spectrum is much less than for the purified native enzyme. In addition, as the light output of the reaction decays, the two components of the spectrum do not decay at the

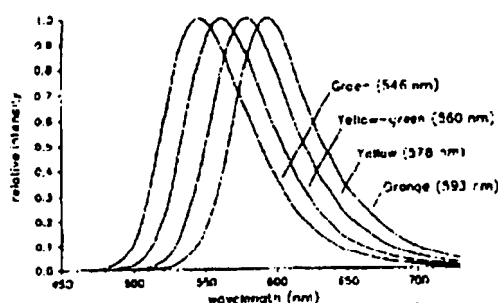


Figure 5. Spectra of bioluminescence emitted from *E. coli* cells containing the click beetle luciferases. The intensity maximum for each spectrum has been normalized.

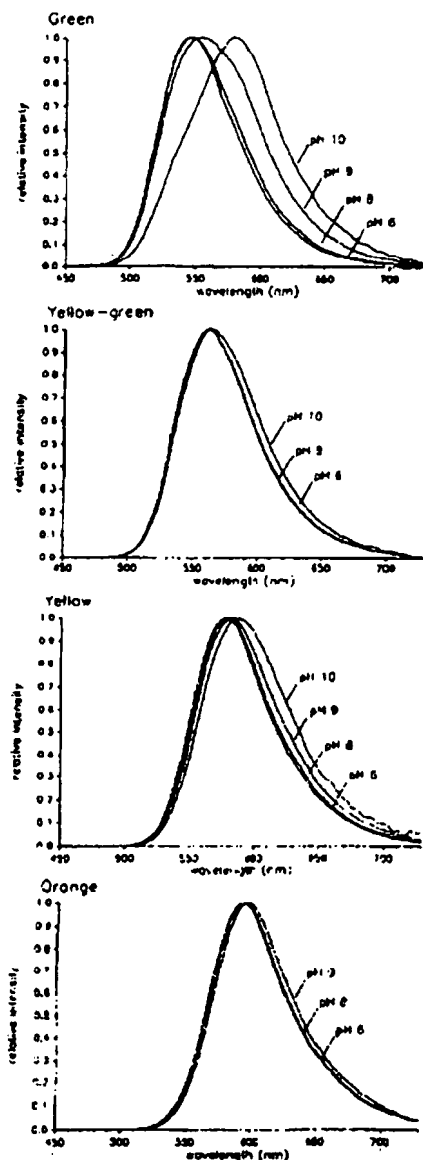


Figure 6. Spectra measured from partially purified lysates of *E. coli* expressing the click beetle luciferases. The intensity maximum for each spectrum has been normalized. The colour emitted by each luciferase at neutral pH is indicated in the corner of each plot. Spectra shown for pH 6.0 and 8.0 were measured in MES/MOPS/Tricine buffer. Spectra shown for pH 9.0 and 10.0 were measured in Tricine/CHES buffer. At pH 8.0, the spectrum measured in MES/MOPS/Tricine was identical to that measured in Tricine/CHES.

same rate. The decay rate of these two components is the same rate in the spectrum of the native luciferase. The spectra of the luciferase produced in *E. coli* are of samples that are only partially purified by the method stated above. It can be shown that the differences between this and the native luciferase are not due to intrinsic differences in the enzymes themselves, but instead arise from the effects of the other components in the bacterial lysate. If the native luciferase is mixed with a lysate prepared from *E. coli* which does not contain a luciferase cDNA clone, the spectrum of the mixture is the same as that of lysates containing the luciferase produced in *E. coli*.

These observations reveal two aspects of the effects of an *E. coli* lysate on the spectrum of firefly luciferase. One is that the lysate contains a component that causes luciferase to resist the effects of pH on its spectrum. The other feature is

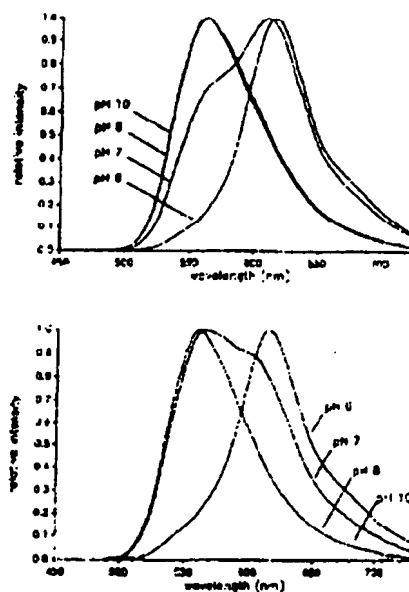


Figure 7. Spectra measured from purified native firefly luciferase (upper plot) and from partially purified lysates of *E. coli* expressing the firefly luciferase (lower plot). The intensity maximum for each spectrum has been normalized. Spectra shown for pH 6.0 and 7.0 were measured in MES/MOPS/Tricine buffer. Spectra shown for pH 8.0 and 10.0 were measured in Tricine/CHES buffer. At pH 8.0, the spectrum measured in MES/MOPS/Tricine had a slightly greater contribution from the red component than that measured in Tricine/CHES.

that firefly luciferase, in the presence of the lysate, is in at least two different forms distinguished both by the colour of light emitted, and by different decay rates of the light output. These two features may indicate a single phenomenon. That is, a factor in the lysates may be stabilizing some of the luciferase molecules both to destabilizing effects of low pH, and to the temporal loss of enzymatic activity. The presence of the bacterial lysate does not appear to affect the spectral distribution of each of the components of the firefly luciferase spectrum, just their relative contribution to the total spectrum. While this effect is most evident when the spectrum is measured at pH 7.0, it is also evident at pHs 6.0 and 8.0. In these cases, however, the differences are slight since the spectrum consists almost entirely of a single component. Extrapolation of these results to the spectra of the click beetle luciferases indicate that their spectral distributions in the pH range of 6 to 8 are probably not affected by the lysate. This is true since the spectrum of these luciferases is apparently a single component in this pH range. But the pH required to shift the spectra to longer wavelength is potentially different than what would be expected for purified enzymes. However, the spectra of the green-emitting click beetle luciferase at pH 9.0 or 10.0, which also consists of two components, did not reveal the nonuniform decay rate evident with the firefly luciferase at pH 7.0.

#### SEQUENCE COMPARISON OF CLICK BEETLE AND FIREFLY LUCIFERASES

Our inability to demonstrate cross-hybridization of their corresponding nucleic acid in Southern blots suggested that a significant degree of evolutionary divergence had occurred between the firefly and click beetle luciferases. Sequence analysis of the click beetle cDNA clones has confirmed this. For a direct comparison with the firefly luciferase, the yellow-green-emitting click beetle luciferase was used since its spectral maximum is at nearly the same wavelength. The cDNA encoding this luciferase contains an open reading frame corresponding to 543 amino acids. This is seven amino acids less than that found with the firefly luciferase cDNA. Alignment of the amino acid sequences, deduced from the cDNA sequences, reveals a 47% identity between the

two luciferases (Fig. 8). The difference in the number of amino acids between the sequences is mostly accounted for by six gaps in the sequence alignment. These gaps are small, being one or two amino acid in length and, for some, their exact position is somewhat arbitrary.

Throughout the alignment there are no regions of especially high sequence similarity. Thus there is no indication of which regions may have been conserved owing to catalytic or structural constraints on the enzymes. An exception to this is in the last three amino acids which, for the firefly luciferase, have been shown to be necessary for translocation into peroxisomes (Gould and Subramani, 1988b, 1988c). Given the close functional similarity of these enzymes, it is almost certain that the click beetle luciferases are also located in peroxisomes. The hydropathy plots of firefly and click beetle luciferase shown some similarities, but overall appear to be quite different (Fig. 9) (Kyte and Doolittle, 1982). The most apparent similarity is the three large hydrophobic regions found in both luciferases. Other regions of hydrophobicity and hydrophilicity can be found in common between the luciferases, but they are largely obscured in as many differences.

In contrast to the low degree of similarity between the firefly and click beetle luciferases, the similarity between the various click beetle luciferases is very high. Between the luciferases which are capable of emitting different colours, the amino acid sequences are from 95% to 99% identical (Wood *et al.*, 1988b). Since the only difference between these luciferases are the amino acid sequences, the determinants of colour must be found in the relatively few differences between the sequences. We have begun to examine exactly which of the amino acids can affect the colour of light, and have found that not all of the differences between the clones are effective. In some cases, the amino acid determinants of colour may be as few as two or three. This work will be presented elsewhere.

#### USE OF CLICK BEETLE LUCIFERASES AS GENETIC REPORTERS

It appears that the click beetle luciferases will have all the advantages of the firefly luciferase in their application as reporters of genetic activity. Advantages such as the sensitivity with which they can be detected, or the ability to detect them



in living cells, are evident. Other features of the luciferases, such as their specific activity, or the linearity of their assay with respect to enzyme concentration, have yet to be established. We are at present improving the expression of these luciferases in *E. coli* to provide a source of enzyme for better characterization. It may be expected with the large difference in sequence between the click beetle and firefly luciferases, that these luciferases also have significant differences in their chemical properties as well. This is supported by the dramatic difference in the response of their spectra to changes in pH. By qualitative observation, temperatures above 40°C or the presence of  $Zn^{2+}$  do not cause changes in the spectra of the click beetle luciferases. As noted above, these conditions will cause the firefly luciferase to emit red light. In fact, the temperature optimum for the click beetle luciferases may be higher than for the firefly luciferase. Other initial experiments suggest that the click beetle luciferases may be more resistant to denaturation by charged detergents, or activation by neutral detergents (Kricka and Deluca, 1982). Collectively, these observations suggest that the activity of the click beetle luciferases may be less sensitive to environmental conditions. However, these conclusions are tentative since they were made from luciferases in the presence of other components of the *E. coli* lysate.

A novel feature of the click beetle clones is the ability to distinguish between them by the colour of light emitted. This may make them particularly useful as genetic reporters where multiple reporters are desirable. Because their respective sequences differ by only a few amino acids, characteristics of their expression in exogenous hosts should also differ little. The differences in the colour of light would normally have no effect on the hosts, but regardless, expression of a luciferase reporter gene is generally done in the absence of the luciferin substrate. Thus there is no luminescent activity until the actual time of the luciferase assay. The spectral distribution of the luciferases are rather broad which would limit the ability to distinguish each luciferase in a mixture if their respective amounts vary widely. The greatest distinction can be made between the green and orange-emitting clones, which should be distinguishable in a luminometer with the use of optical cut-off filters. From calculations based on the overlap of their spectra alone, and assuming a coefficient of variation of 4% in the

assay of luminescence, this method should allow the detection of one of the luciferases in the presence of a 25-fold excess of the other. Since the colours of these luciferases are not easily altered by pH or temperature, it should be possible to distinguish these luciferases *in vivo* as well as *in vitro*. This type of dual reporter gene system would allow one to monitor different promoters within a single host, or to follow different populations of cells simultaneously, each labelled with a different luciferase. The structural similarity of the luciferases increases confidence that differential effects noted in an experiment are properties of the system being observed, and are not artefacts due to individual peculiarities of the reporter genes themselves.

## SUMMARY

Firefly luciferase has been used as a tool of scientific investigation for over two decades because of the high sensitivity with which its enzymatic activity can be assayed. With the advent of techniques in nucleic acid manipulations, it has found its newest area of application as a reporter of genetic activity within living cells. In addition to high sensitivity, its assay is rapid and does not require complex procedures or precautions. In comparison to the CAT assay, firefly luciferase has been shown to be well suited as a genetic reporter. But, whereas previously firefly luciferase was the epitome of beetle luciferases because of its availability, cloning techniques have made feasible the study of other luciferases of this type. Some of these luciferases may have additional features enhancing their use as reporters, or in other applications. Our recent cloning of several luciferases from a bioluminescent click beetle substantiates this possibility. These luciferases are unique in the ability to produce bioluminescence of several different colours. In addition, the sequence of these luciferases is considerably different from that of the firefly luciferase, suggesting that other chemical properties of these enzymes will be different. One area where such differences are apparent is in the response of the bioluminescence spectra to changes in pH. We are currently investigating other properties of these new luciferases to better understand their general nature and to determine their suitability in applications.



## REFERENCES

- Alter, S. C. and DeLuca, M. (1986). The sulfhydryls of firefly luciferase are not essential for activity. *Biochemistry*, **25**, 1599-1605.
- Biggley, W. H., Lloyd, J. E. and Seliger, H. H. (1967). The spectral distribution of firefly light. II. *J. Gen. Physiol.*, **50**, 1681-1692.
- Crenshaw, E. D. and Rosenfeld, M. G. (1988). (manuscript in preparation).
- Crowman, R. A. (1981). *The Biology of the Coleoptera*. Academic Press, London.
- de Boer, H. A., Comstock, L. J. and Vasser, M. (1983). The lac promoter: a functional hybrid derived from the *irp* and *lar* promoters. *Proc. Nat. Acad. Sci. USA*, **80**, 21-25.
- DeLuca, M. and McElroy, W. D. (1978). Firefly luciferase. In *Methods in Enzymology: Bioluminescence and Chemiluminescence*, Vol. LVII, DeLuca, M. (Ed.), Academic Press, Orlando, pp. 3-14.
- DeLuca, M., Leonard, N. J., Gates, M. J. and McElroy, W. D. (1973). The role of 1,N<sup>6</sup>-ethenoadenosine triphosphate and 1,N<sup>6</sup>-ethenoadenosine monophosphate in firefly luminescence. *Proc. Nat. Acad. Sci. USA*, **70**, 1664-1666.
- de Wet, J. R., Wood, K. V., Helinski, D. R. and DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proc. Nat. Acad. Sci. USA*, **82**, 7870-7874.
- de Wet, J. R., Wood, K. V., Helinski, D. R. and DeLuca, M. (1986). Cloning firefly luciferase. In *Methods in Enzymology: Bioluminescence and Chemiluminescence*, Part B, vol. 133, DeLuca, M. and McElroy, W. D. (Eds), Academic Press, London, pp. 3-14.
- de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. and Subramani, S. (1987). The firefly luciferase gene: structure and expression in mammalian cells. *Molec. Cell. Biol.*, **7**, 725-737.
- DiLitta, A. G., Hope, D. A., Chen, H., Trumbauer, M., Schwarz, R. J. and Smith, R. G. (1988). Utility of firefly luciferase as a reporter gene for promoter activity in transgenic mice. *Nucl. Acids Res.*, **16**, 4159.
- Gould, S. J. and Subramani, S. (1988a). Firefly luciferase as a tool in molecular and cell biology. *Anal. Biochem.* (in press).
- Gould, S. J. and Subramani, S. (1988b). *J. Cell. Biol.* (in press).
- Gould, S. J. and Subramani, S. (1988c). (manuscript in preparation).
- Howard, P. K., Ahern, K. G. and Firtel, R. A. (1988). Establishment of a transient expression system for *drosophila* discoidium. *Nucl. Acids Res.*, **16**, 2613-2623.
- Keller, G. A., Gould, S., DeLuca, M. and Subramani, S. (1987). Firefly luciferase is targeted to peroxisomes in mammalian cells. *Proc. Nat. Acad. Sci. USA*, **82**, 3264-3268.
- Kricha, L. J. and DeLuca, M. (1982). Effect of solvents on the catalytic activity of firefly luciferase. *Arch. Biochem. Biophys.*, **217**, 674-681.
- Kyle, J. and Donkille, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.*, **157**, 105-132.
- Lloyd, J. E. (1978). Insect bioluminescence. In *Bioluminescence in action*, Herring, P. J. (Ed.), Academic Press, London, pp. 241-272.
- McElroy, W. D. and Seliger, H. H. (1960). Spectral emission and quantum yield of firefly bioluminescence. *Arch. Biochem. Biophys.*, **88**, 136.
- McElroy, W. D. and Seliger, H. H. (1966). The colors of bioluminescence: role of enzyme and substrate structure. In *Molecular Architecture in Cell Physiology*, Hayashi T. and Szere-Nyergyi A. (Eds), Prentice-Hall, N.J. pp. 63-80.
- Ow, D. W., Wood, K. V., DeLuca, M., de Wet, J. R., Hellaski, D. R. and Howell, S. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, **236**, 856-859.
- Seliger, H. H. and McElroy, W. D. (1964). The colors of firefly bioluminescence: enzyme configuration and species specificity. *Proc. Nat. Acad. Sci. USA*, **52**, 75-81.
- Seliger, H. H., Buck, J. B., Fastie, W. G. and McElroy, W. D. (1964). The spectral distribution of firefly light. *J. of Gen. Physiol.*, **48**, 95-104.
- Short, J. M., Fernandez, J. M., Surge, J. A. and Huse, W. D. (1988). Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucl. Acid. Res.*, **16**, 7585-7601.
- Wienhausen, G. and DeLuca, M. (1985). Luciferases from different species of fireflies are antigenically similar. *Photochem. Photobiol.*, **42**, 609-611.
- Wood, K. V. and DeLuca, M. (1987). Photographic detection of luminescence in *Escherichia coli* containing the gene for firefly luciferase. *Anal. Biochem.*, **161**, 501-507.
- Wood, K. V., de Wet, J. R., Dewji, N. and DeLuca, M. (1988). Synthesis of active firefly luciferase by *in vitro* translation of RNA obtained from adult lanterns. *Biochem. Biophys. Res. Commun.*, **124**, 592-596.
- Wood, K. V., Lam, Y. A. and McElroy, W. D. (1988a). (manuscript in preparation).
- Wood, K. V., Lam, Y. A. and McElroy, W. D. (1988b). Bioluminescent click beetles revisited. *Biolum. Chemlum.* (in press).

## ***Luc* Genes: Introduction of Colour into Bioluminescence Assays**

**Keith V. Wood**

Department of Chemistry, M-001, University of California, San Diego, La Jolla, CA 92093, USA

Luminescence assays are generally based on measurements of light intensity alone. Inclusion of colour as an additional parameter of the assay could increase the information content. Colour variation in luminescence is particularly prevalent among beetle luciferases. To study the relationship between enzyme structure and colour, luciferases from a Jamaican click beetle were examined as a model system. These luciferases emit light ranging from green to orange, though their amino acid sequences differ by less than 5%. Through mutation of their respective cDNA clones, the amino acids responsible for the colour variation were identified. These specific amino acids are few, and they act upon colour independently with respect to the enzyme structure. Analysis of their effects indicates that the potential for colour variation among beetle luciferases is greater than is evident among the click beetle luciferase. Because of the subtle changes of enzyme structure that effect colour, luciferases that emit different colours may be useful as paired genetic reporters. They should interact equivalently with the intracellular environment of a host, but could be distinguished by colour in their assay. Such paired reporters could be used to observed simultaneous events, or to provide internal control for luminescence measurements.

**Keywords:** Firefly luciferase; click beetle luciferases; reporter genes; colour variation

### **INTRODUCTION**

The production of light by enzymatic catalysis offers unique opportunities for probing biochemical processes. This is because of the high energy density of photons and their unusual presence in a biochemical milieu. In applications of bioluminescence, the chemistries of luminous bacteria and beetles have been dominant. Although the mechanisms of these two systems are entirely different, they both are amenable to manipulations: the enzymes are reasonably stable and easily purified, the luciferins are available by chemical synthesis, and the other substrates are readily obtainable in pure form.

Luminescence as a biochemical guage is based

on correlating light with a limiting component of the enzymatic reaction—changes in concentration of the limiting component cause proportionate variation in light emission. Initial applications of luminescence used cofactors as the limiting components. For example, firefly luciferase has been widely used to measure ATP. Similarly, bacterial luciferase coupled to an oxidoreductase has been used to measure NADH. With coupling to other enzymes, these luciferases have also been used to measure other biochemical molecules (McElroy and DeLuca, 1983).

Recently, a new class of luminescence applications has arisen where the enzyme is the limiting component. In these applications, light emission is linked to events associated with gene regulation

and protein metabolism. This was made possible with the cloning of genes that code the luciferases (Cohn *et al.*, 1983; de Wet *et al.*, 1985). These genes can be introduced into living cells, or reconstituted enzyme system, so that the synthesis of luciferase is contingent upon the kinetics of gene expression. With excess substrates, luminescence is proportional to the concentration of newly synthesized enzyme.

Because of the prevalence of research in molecular genetics, applications of bioluminescence are most auspicious in this area. In eukaryotic systems especially, the use of firefly luciferase has been notable. This monomeric enzyme, evolved in a eukaryotic host, requires no post-translational modifications for its catalytic activity. Under optimal conditions, it catalyses production of yellow-green light with exceptional efficiency (McElroy and DeLuca, 1985; Seliger and McElroy, 1960).

The general suitability of this luciferase as a genetic reporter has made it useful in a variety of experimental designs. Most commonly it has been used in examining the DNA structure of genetic regulatory elements (Economou *et al.*, 1989; Hudson *et al.*, 1989; van Zonneveld *et al.*, 1988). Some studies have used this luciferase to investigate other proteins that influence gene transcription (Mellon *et al.*, 1989; Waterman *et al.*, 1988). Also studied have been effects of mRNA structure on protein synthesis (Malone *et al.*, 1989; Baughman and Howell, 1988), and relative rates of intracellular protein recycling (Nguyen *et al.*, 1989). The firefly luciferase has in some instances been used to delineate genetic events in multicellular organisms (Rodriguez *et al.*, 1989; Ow *et al.*, 1986).

Common to these luminescence applications is that measurements are made of light intensity alone. However, this is only one mode by which light can carry information. Another prominent property of light is its spectral distribution, i.e. the colour of light. If this property could be used in addition to intensity, it could add another dimension to the information transmitted by the luciferases. Each luciferase elicits a characteristic spectral distribution. Even within the distinct groups of beetle or bacterial luciferases there is variation of colour. Since the substrates within these groups do not differ, the colour variation must be due to differences in enzyme structures.

Colour variation is especially prominent among the beetle luciferases. A spectacular example of this variation occurs in a tropical click beetle from Jamaica, *Pyrophorus plagiophthalmus*. The beetle

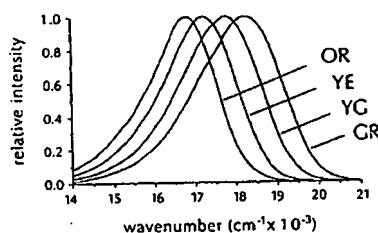
has two sets of light organs, a pair on the dorsal surface of the prothorax, and a single organ in a ventral cleft of the abdomen. Generally the dorsal pair emits green light, and the ventral organ emits yellow light. Hence, this is an unusual example of an organism that emits two different colours of light. Even more unusual is that variation in colour occurs between individuals of the population. The dorsal organ varies in colour from green to yellow-green, and the ventral organ varies from green to orange (Biggley *et al.*, 1967).

Because of the wide range of colours found in this single species, it was chosen as a model of colour variation among beetle luciferases. Research has begun to investigate the relationship between enzyme structure and the colour of luminescence. Results of the ongoing project have revealed some general aspect of this relationship. Substantial changes in colour can result from substitutions of single amino acids in the primary structures of the enzymes. These substitutions can occur at several different positions, and the effect of different substitutions act independently. A quantitative analysis of several substitutions has indicated that the potential for colour variation in beetle luciferases is much greater than the range of colours found in this particular beetle species. These results foretell the feasibility of using colour as an additional parameter in luminescence assays.

#### COLOUR VARIATION IN *P. PLAGIOPHTHALMUS*

To study the luciferases of the Jamaican click beetle, the cloning techniques previously employed to clone the firefly luciferase were used (de Wet *et al.*, 1985). The luciferases of the ventral light organ were chosen for initial study because of their wider range of colour variation. Screening a cDNA library made from this organ, both for antigenic epitopes and for luminescence activity, resulted in 11 clones with complete coding regions. When expressed in *E. coli*, these clones can produce sufficient bioluminescence to be easily visible.

The clones are of four types determined by the colour of light elicited: green, yellow-green, yellow, and orange (Fig. 1). Among the seven clones that produce yellow light, or the three that produce orange light, the spectra are indistinguishable. Only one green and one yellow-green light-producing clone were obtained. As determined by the positions of the peak intensities, the range of colours



**Figure 1.** Luminescence spectra of the click beetle luciferases: GR, green; YG, yellow-green; YE, yellow; OR, orange. Intensity maxima are normalized for comparison

produced by the clones is the same as was measured from living beetles (Biggley *et al.*, 1967). However, there are colours displayed by living beetles that are not represented by the clones. Some of these colours may be the result of heterozygous beetles expressing a mixture of luciferases. Of course, the genes of some colours may not have been cloned yet.

The amino acid sequences of the click beetle luciferases were determined from the nucleic acid sequences of the clones. In paired comparisons these amino acid sequences are 96% to 99% identical, differing by 26 to 3 amino acids respectively. The comparisons reveal the genealogy of these enzymes, which shows that they have evolved in the order of their colours. Thus, luciferases of similar colour are of more similar sequence. Also, the most recently evolved luciferase emits orange light; the green-emitting luciferase is the oldest. So the sequences of the luciferases emitting orange and yellow are more similar than those of the green and yellow-green.

As noted above, the variation in colour must lie within differences of the enzyme structures. Since there is no evidence of post-translational modification in the luciferases, the differences should be evident within the amino acid sequences. To determine which (and how many) of the amino acids affect colour, mutants of the luciferases were made by modifying their cDNA clones. The mutants were made by two methods. Many were made simply by exchanging restriction fragments between the clones. The resulting new luciferases were named rearrangement hybrids. Other mutants were made using synthetic oligonucleotides for site-specific changes.

Because the rearrangement hybrids were made by swapping segments of genetic code, they often contain multiple amino acid substitutions. Nota-

tion to describe these substitutions is, for example,  $R_{223}, L_{238} \rightarrow E, V$ . This depicts the substitution of arginine at position 223 and leucine at position 238 for glutamine and valine respectively. The resulting changes in the colour are reported in wave numbers instead of wavelength since wavenumbers are proportional to energy, an additive quantity. In the example, the substitutions cause a colour shift of  $-520 \text{ cm}^{-1}$ , from  $17,760 \text{ cm}^{-1}$  of the parent luciferase to  $17,240 \text{ cm}^{-1}$  of the progeny.

The results of studying several mutants show that colour differences among the yellow-green-, yellow-, and orange-emitting luciferases are due predominantly to three amino acid substitutions. The colour difference between the yellow- and orange-emitting luciferases is due entirely to  $S_{247} \rightarrow G$ . Approximately 90% of the colour difference between the yellow-green- and yellow-emitting luciferases is due to two substitutions,  $R_{223} \rightarrow E$  and  $L_{238} \rightarrow V$ ; the colour shift caused by  $L_{238} \rightarrow V$  is about 1.3-fold greater than that of  $R_{223} \rightarrow E$ . The remaining 10% of colour difference is due to one or more substitutions of  $L_{41}, D_{226}, V_{282}, I_{283}, V_{323}, V_{389} \rightarrow I, E, IV, I, I$ . Likewise, the amino acids affecting colour of the green-emitting luciferase have not yet been precisely determined because of the large number of sequence differences between this and the other luciferases.

From other mutants it is evident that the substitutions that affect colour do so regardless of the parent luciferase. For example, the  $S_{247} \rightarrow G$  substitution, which causes the yellow-emitting luciferase to produce orange light, can be applied to the yellow-green-emitting luciferase. This causes an analogous shift of colour to yellow. Similarly,  $E_{223} \rightarrow R$ , which partially effects the shift from yellow to yellow-green, produces an analogous shift towards green when applied to the orange-emitting luciferase. Apparently, colour differences evolved in the click beetle luciferases by the cumulative effects of individual substitutions. That is, the change from yellow-green to orange requires the combined action of  $R_{223} \rightarrow E$ ,  $L_{238} \rightarrow V$ , and  $S_{247} \rightarrow G$  (with small contributions from other substitutions). Presumably, this general scheme applied also to the green-emitting luciferase.

#### INDEPENDENCE OF SUBSTITUTIONS AFFECTING COLOUR

When amino acid substitutions in an enzyme act independently, the action of one substitution

should not affect the action of the other. As a corollary, the individual effects of the substitutions should be additive. In the opposite situation, extreme cooperativity, the action of one substitution is fully dependent on the other. Among the click beetle luciferases, the cumulative action of the substitutions that affect colour rules out extreme cooperativity. Superficially, however, the substitutions do not appear to act entirely independently.

For example, the rearrangement hybrid with substitutions  $E_{223}, V_{238} \rightarrow R, L$  causes the colour of the yellow-emitting luciferase to shift  $490 \text{ cm}^{-1}$ . However, the combined effects of  $E_{223} \rightarrow R$  and  $V_{238} \rightarrow L$  applied to the yellow-emitting luciferase is a shift of  $420 \text{ cm}^{-1}$ . Thus, the individual affects appear to be 14% less than their combined affect. As another example, the substitutions  $S_{247} \rightarrow G$  applied to the yellow-emitting luciferase causes a shift of  $-430 \text{ cm}^{-1}$ . However, when the same substitution is applied to the yellow-green-emitting luciferase, the resulting shift is  $-580 \text{ cm}^{-1}$ , 35% greater.

Close examination of the shifts caused by each type of substitution has revealed a consistent relationship with regard to the colour of the parent luciferase. Specifically, substitutions have a greater affect on colour when applied to luciferases of greater wavenumber. This effect is shown quantitatively in a plot of shift magnitude vs shift position (Fig. 2). The magnitude is simply the absolute value of the difference in positions of spectral maxima between the parent and progeny luciferases, i.e. the shift without regard to sign. The positions of the shift was taken as the average of the position of spectral maxima for the parent and progeny luciferase. This was chosen instead of the position for the parent luciferase since, in theoretical considerations, the distinction between parent and progeny is arbitrary. Thus, by using the average as a measure of position, the choice of parent or progeny is mute.

It is apparent from the plot that the different types of substitutions define a converging set of trends. The substitutions (or groups of substitutions), whose resulting shift are shown as filled symbols in Fig. 2, represent independent sets. The grouped substitutions, whose resulting shifts are shown as open symbols, include within themselves substitutions of the independent sets. Although these shifts do not result from independent substitutions, they are derived from independent measurements of rearrangement hybrids. Thus, they provide additional evidence for the trends.

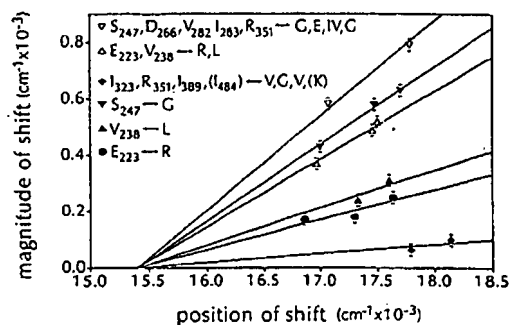


Figure 2. Relationship of shift magnitude to the position of the shift. Key to amino acid substitutions is shown on the left. Filled symbols show shifts caused by independent substitutions; open symbols show shifts resulting from combined sets of substitutions. Bars above and below each data point indicate one standard deviation of experimental error. See text for criteria of the axes and description of interpolated lines

Using linear least-squares analysis, a set of lines was determined for the shifts of each set of substitutions. Most of these lines cross the abscissa near  $15,500 \text{ cm}^{-1}$ . The lines drawn in Fig. 2 are an interpretation of the data. These lines converge to a single point on the abscissa. The inference is that there exists a unique minimum for the position of the luminescence spectrum. Effects of the substitutions are not able to exceed that minimum. That is, there is a limit to how red the luminescence can be and no alterations to the luciferase structure can result in light that is more red.

This seems reasonable with regard to the structure of the light-emitting molecule. Beetle luciferases catalyse light production by combining ATP and luciferin to form luciferyl-AMP, which is then oxidized to oxyluciferin. The oxyluciferin is formed in an electronically excited state, and a photon is generated upon its transition to the ground state. The colour of luminescence is determined by the energy difference between the ground and excited states. This difference is influenced by interactions of the oxyluciferin with the enzyme structure. Therefore, colour can be affected by substitutions of specific amino acids. However, without changes to the bonding structure of oxyluciferin, there should exist a minimum possible energy difference between the ground and excited states. This minimum would impose a minimum of the energy of the emitted photon, i.e. the colour of light.

The lines in Fig. 2 were drawn to converge at  $15,400 \text{ cm}^{-1}$ . This value is the position of maximal

intensity for the luminescence generated by oxidation of luciferyl-AMP in aqueous buffer without enzyme (White *et al.*, 1971). It is the lowest wavenumber value measured for the luminescence of oxyluciferin under any conditions, with or without enzyme. It is known that the colour of luminescence elicited from non-enzymatic oxidation is dependent on the polarity of the solvent. Under less polar conditions, such as in DMSO, the spectrum of luminescence shifts to greater wavenumbers (White *et al.*, 1971). Water is a highly polar medium and may provide conditions for luminescence of minimum energy.

Drawing the trends of spectral shifts as lines converging at  $15,400\text{ cm}^{-1}$  is in good agreement with the data. Each of the data points is near its respective line within one standard deviation of experimental error. Thus, the empirical data is in accord with the hypothesis of an energy minimum for the luminescence. Since this minimum was measured from a non-enzymatic reaction, it endorses the belief that this minimum is determined entirely by physical properties of oxyluciferin. Hence, where the lines converge in Fig. 2 should be independent of structural features particular to the click beetle luciferases. Moreover, colour changes caused by amino acid substitutions in any beetle luciferase should exhibit trends that converge to this same minimum value.

The apparent lack of independence noted above between colour shifts is evident in the slopes of the lines in Fig. 2. If the shifts had displayed the sense of independence describe at the beginning of this section, then the slopes would have to be zero. That is, regardless of the colour of the parent luciferase, the substitutions would cause the same magnitudes of shift. Yet, slopes of zero would imply no minimum to the energy of luminescence. Thus, the interdependence between the effects of substitutions imposed by the slopes is due to physical limitations of the substrate.

Within this constraint, however, the effects of the substitutions behave completely additively. For example, calculated from the lines of Fig. 2,  $E_{223} \rightarrow R$  applied to the yellow-emitting luciferase would result in a shift of  $204\text{ cm}^{-1}$ . Applying  $V_{238} \rightarrow L$  to the resulting hypothetical luciferase would result in an additional  $288\text{ cm}^{-1}$ . The sum of these shifts,  $492\text{ cm}^{-1}$ , is only 1% less than the expected value of  $E_{223}, V_{238} \rightarrow R, L$  applied to the yellow-emitting luciferase,  $498\text{ cm}^{-1}$ . Similarly, a shift caused by  $S_{247}, D_{266}, V_{282}, I_{283}, R_{351} \rightarrow G, E, IV, G$  is equal to the sum of its component substitutions. Because the

trends of Fig. 2 are described by lines, additive relationships demonstrated for one luciferase are the same when applied to other luciferases. Also, it follows that the order in which the substitutions are considered is unimportant.

Therefore, respective to the slopes of the lines, the effects of the substitutions act fully independently. In equivalent terms, the substitutions are independent with regard to their action on the substrate; the apparent interdependence of colour shifts can be attributed to properties of oxyluciferin. There is no evidence of dependent relationships mediated by the structures of the enzymes. That is, there are unlikely to be any interactions within the enzyme structures between the amino acids at the positions of the substitutions.

#### POTENTIAL FOR COLOUR VARIATION AMONG BEETLE LUCIFERASES

The colours of luminescence emitted by the Jamaican click beetle define nearly the full range of colours found in all luminous beetles (Lall *et al.*, 1980). Yet, the trends depicted in Fig. 2 suggest a potential for colour variation in beetle luciferases that is much greater. The range spanned by the click beetle luciferases is  $1400\text{ cm}^{-1}$ . If the lower end of this range were extended to its theoretical limit, the full range would double to  $2800\text{ cm}^{-1}$ . Furthermore, there is no indication in Fig. 2 of an upper limit to the possible range. Certainly an upper limit exists owing to conservation of energy in the luminescent reaction. However, it is unknown what further considerations could impose a more strict upper limit.

The ability of beetle luciferases to support redder colours of luminescence is evident in the luminescence of the firefly luciferase (*P. pyralis*). Though this enzyme normally emits yellow-green light, under several conditions it emits red light of  $16,160\text{ cm}^{-1}$  (McElroy and DeLuca, 1985). Some of these conditions are pH below 7, temperature above  $30^\circ\text{C}$ , and the presence of heavy metals such as  $\text{Hg}^{2+}$ . Chemical modification to the enzyme can also result in red luminescence (Alter and DeLuca, 1986). This red colour extends the range of enzymatic luminescence by 50% over that of the click beetle luciferases alone.

Nature has also provided one known example of red beetle luminescence in a rare species called *Phrixothrix*. This larviform beetle of South America has two rows of light organs that emit yellow-

green light, and a pair near the head that emit red light. Although the spectra of this red has not been measured, it is of much lower wavenumber than the orange of the click beetle. Thus, red luminescence is also possible through the mechanism of natural evolution. Since post-translational modifications that affect colour are not found in either the firefly or click beetle luciferases, the evolution of *Phrixothrix* luciferase is probably also mediated by modifications to the amino acid sequence.

So why do virtually all species of luminous beetles emit light in the limited range of green to yellow? The above discussion indicates that the enzyme is capable of supporting a much larger range. The reason may be in the 'motive' of beetle luminescence. The system has evolved to maximize communication between beetles, i.e. to maximize visibility. An essential aspect of this is the way in which colour interacts with the environment. For example, green is the colour of maximum reflectivity for foliage. Also, measurements of ambient light at dusk in a foliated area reveal a minimum near yellow (Seliger *et al.*, 1982b). The colours of beetle luminescence may be partially dictated by these environmental parameters, depending on the behavioural characteristics of the species. Evidence for this has been documented for firefly luminescence (Seliger *et al.*, 1982a,b).

However, the needs of beetle communication do not necessarily equal the needs of genetic research. In applications to utilize luciferase of different colours, a wide range would be more useful. Further study of colour variation in beetle luciferases should allow development of synthetically modified enzymes that elicit colours not found in nature. This could accord access to the full colour potential of this luciferase system. The lesson of the click beetle luciferases is that such modifications may encompass only substitutions of independently acting amino acids. Moreover, this natural example shows that there may be many candidates for such substitutions.

This follows from the relatively recent evolutionary history that brought about the different colours of the click beetle. Evolution is unable plan in its course; it can only operate by selection of randomly provided mutations. Yet, in the 26 amino acids that distinguish the four click beetle luciferases, more than four affect colour. Thus, by trial-and-error evolution, acceptable candidates for colour variation were rapidly found. It can be inferred that little of the mutagenic potential of these enzymes was tested by this process from comparison with the

firefly luciferase. The luciferases from these two beetle species differ in amino acid sequence by 51%. Thus, amino acid substitutions are potentially acceptable at 275 positions. For each position, several different amino acids may suffice. In brief, the beetle luciferases appear to hold much potential for modification and much potential for variation in colour of luminescence.

## APPLICATION OF COLOUR IN LUMINESCENCE ASSAYS

In luminescence assays, light intensity is the signal that conveys information of biochemical events. With the use of two luciferases that emit light of different colours, it should be possible to provide two signals simultaneously. Beetle luciferases offer the potential of providing such a two-coloured system. This probably would be most useful in applications associated with molecular genetics. This is because applications utilizing two colours would generally depend on the enzymes as the limiting components of the assay. The concentration of the enzymes must be the source of the signals since the distinction of colour lies in the enzyme structures. Assays based on a substrate as the limiting component, ATP for example, would not benefit because the enzymes of both colours would generate signals dependent on the same condition. Thus, additional information could not be gained with the second colour.

The obvious use of two colours would be for simultaneous detection of two different events. This would be especially useful when the events are coordinated. An example could be the transcriptional activities of promoters regulated by a common mechanism. Moreover, the promoters need not be of a single host, such as with regulation mediating symbiotic or parasitic relationships. The luciferases also need not be used for quantitative measurements, but merely as markers for two populations. For instance, populations of a colony-forming organism could be identified visually by their colour of luminescence. The genes coding the luciferases may also offer a method of detecting genetic recombination events, depending on the positions of the colour-determining nucleic acid substitutions.

Another general use of two colours would be for provisions of an internal control in luminescence measurements. Precision in genetic measurements can be important, especially in eukaryotic hosts

where differences of two- to three-fold are significant. Internal controls are often needed to compensate for uncontrollable variables. A common example is in experiments where DNA is introduced into cells for measurements of transient gene expression. To compensate for variation in the efficiency of transfection, a second genetic reporter is sometimes included (Dirks *et al.*, 1989; Day and Maurer, 1989).

After a gene is introduced into a cell, there are other potential variables of gene expression, such as rates of translation and protein stability. These variables that concern the behaviour of a reporter within a cell are more difficult to compensate for by using a second reporter. The problem is that different reporters can behave quite differently in a common host. For instance, comparisons of luciferase to another commonly used reporter, chloramphenicol acetyl transferase (CAT), reveal substantially different kinetics of expression (Maxwell and Maxwell, 1988). As expected, the structures of these dissimilar proteins interact differently with the complex metabolism of the host.

The ideal solution would be to use reporters whose structures are identical, yet could be distinguished in their assay. Herin could be the major advantage of beetle luciferases. Since only a few amino acid substitutions are needed to alter colour, the overall structures of the reporters could be virtually identical. Hence, there would be little to allow discrimination of their interactions with an experimental host. That is, the host could not differentiate one reporter from the other. Upon assay, however, distinction between the reporters would be made by their colours of luminescence. The similarity between the luciferases would be especially prominent if the distinguishing amino acids were internal to the protein structures. This may be the case since the amino acids that affect colour are likely to be close to the luciferin binding site.

Normally, experimental controls are implemented in genetic experiments through comparisons of a test population with a control population. Inclusion of an internal control would be most useful when inter-experimental variation is large, or replica experiments are difficult to obtain. For instance, replica populations could be difficult to achieve when the experimental host is not derived from a stable clonal source. In this circumstance, comparisons between test and control populations would be difficult. However, an internal control would allow for simultaneous compar-

isons of a test and control within an experiment. The closely matched structures of beetle luciferases could provide a means for internal control in experiments that utilize reporters. The light intensity of one colour would serve as the test signal, and the other the control signal. An example in which this may be useful is in measurements of transgenic organisms. Even though the hosts in these experiments may come from a clonal stock, activity of exogenous genes inserted into their chromosomes can be strongly position-dependent.

An especially promising attribute of beetle luciferases as genetic reporters is the ability to detect their activity from within living cells. Two substrates of the luminescent reaction, ATP and  $O_2$ , are available in the cellular interior. The third substrate, luciferin, can gain access to the interior by diffusion through the membrane. Thus, in cells expressing luciferase, a luminescent signal can be generated for external detection. Because photons are created at the instant of catalysis and do not accumulate, the signal is a 'real-time' indicator of the intracellular luciferase concentration. However, quantitative measurements in living cells can be obscured by several variables. For instance, the enzyme may not be the limiting component of the assay in the intracellular environment since the availability of the other substrates could be limiting. Thus, changes in light intensity could reflect variations in any of the components.

Other factors, such as those described previously, also may conceivably affect intracellular luminescence. Use of two beetle luciferases to provide internal control could compensate for these influences. The test signal of one luciferase would be coupled to the gene of interest, and the control signal coupled to a reference gene. A suitable reference gene could be one of constitutive activity, a so-called 'housekeeping' gene. Measurements would be made not of absolute light intensity of the reporter signal, but by the relative intensity of the test signal compared with the control signal. Conditions that would affect both signals, such as changes in concentration of accessible internal substrates, could therefore be compensated for. By this method of measurement, indications of genetic regulation could be made with direct reference to the baseline genetic activity of the host.

In detail and variation, there doubtless are many ways in which beetle luciferases of different colours could be useful. The general suitability of a single luciferase as a genetic reporter is already amply demonstrated. Since introduction of this applica-



tion, each year has borne increased numbers of citations of its use in the scientific literature. The additional potential offered by luciferases that emit different colours lies in the subtlety of their structural differences. Such a matched pair of reporters, with the sensitivity and versatility of the luciferases, has not been manifested by other systems. Thus, the potential capabilities of beetle luciferases may not only improve current methods of assay, but in addition may endorse new methods. The possibilities presented here are based on our recent knowledge of the beetle luciferases, but further research will be needed to test the limits of these possibilities.

## REFERENCES

- Alter, S. C. and DeLuca, M. (1986). The sulfhydryls of firefly luciferase are not essential for activity. *Biochem.*, **25**, 1599-1605.
- Baughman, G. and Howell, S. H. (1988). Cauliflower mosaic virus 35S RNA leader region inhibits translation of downstream genes. *Virology*, **167**, 125-135.
- Biggley, W. H., Lloyd, J. E. and Seliger, H. H. (1967). The spectral distribution of firefly light, II. *J. Gen. Phys.*, **50**, 1681-1691.
- Cohn, D. H., Ogden, R. C., Abelson, J. M., Baldwin, T. O., Neelson, K. H., Simon, M. I. and Mileham, A. J. (1983). Cloning of the *Vibrio harveyi* luciferase genes: use of a synthetic oligonucleotide probe. *Proc. Natl. Acad. Sci. USA*, **80**, 120-123.
- Day, R. N. and Maurer, R. A. (1989). The distal enhancer region of the rat prolactin gene contains elements conferring response to multiple hormones. *Molec. Endo.*, **3**, 3-9.
- de Wet, J. R., Wood, K. V., Helinski, D. R. and DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **82**, 7870-7873.
- Dirks, W., Mittnacht, S., Rentrop, M. and Hauser, H. (1989). Isolation and functional characterization of the murine interferon-beta 1 promoter. *J. Interferon Res.*, **9**, 125-133.
- Economou, J. S., Rhoades, K., Essner, R., McBride, W. H., Gasson, J. C. and Morton, D. L. (1989). Genetic analysis of the human tumor necrosis factor alpha/cachectin promoter region in a macrophage cell line. *J. Expt. Med.*, **170**, 321-326.
- Hudson, L. G., Santon, J. B. and Gill, G. N. (1989). Regulation of epidermal growth factor receptor gene expression. *Molec. Endo.*, **3**, 400-408.
- Lall, A. B., Seliger, H. H., Biggley, W. H. and Lloyd, J. E. (1980). Ecology of colors of firefly bioluminescence. *Science*, **210**, 560-562.
- McElroy, W. D. and DeLuca, M. (1983). Firefly and bacterial luminescence: basic science and applications. *J. Appl. Biochem.*, **5**, 197-209.
- McElroy, W. D. and DeLuca, M. (1985). Firefly luminescence. In *Clinical and Biochemical Analysis*, Vol. 16: *Chemiluminescence*, Burr, J. G. (Ed.), Marcel Dekker, Inc., New York, pp. 387-400.
- Malone, R. W., Felgner, P. L. and Verma, I. M. (1989). Cationic liposome-mediated RNA transfection. *Proc. Natl. Acad. Sci. USA*, **86**, 6077-6081.
- Maxwell, I. H. and Maxwell, F. (1988). Electroporation of mammalian cells with a firefly luciferase expression plasmid: kinetics of transient expression differ markedly among cell types. *DNA*, **7**, 557-562.
- Mellon, P. L., Clegg, C. H., Correll, L. A. and McKnight, G. S. (1989). Regulation of transcription by cyclic AMP-dependent protein kinase. *Proc. Natl. Acad. Sci. USA*, **86**, 4487-4891.
- Nguyen, V. T., Morange, M. and Olivier, B. (1989). Protein denaturation during heat shock and related stress. *J. Biol. Chem.*, **264**, 10487-10492.
- Ow, D. W., Wood, K. V., DeLuca, M., de Wet, J. R., Helinski, D. R. and Howell, S. H. (1986). Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, **234**, 856-859.
- Rodriguez, D., Rodriguez, J.-R., Rodriguez, J. F., Trauber, D. and Esteban, M. (1989). Highly attenuated vaccinia virus mutants for the generation of safe recombinant viruses. *Proc. Natl. Acad. Sci. USA*, **86**, 1287-1291.
- Seliger, H. H. and McElroy, W. D. (1960). Spectral emission and quantum yield of firefly bioluminescence. *Arch. Biochem. Biophys.*, **88**, 136-141.
- Seliger, H. H., Lall, A. B., Lloyd, J. E. and Biggley, W. H. (1982a). The colors of firefly bioluminescence—I. Optimization model. *Photochem. Photobiol.*, **36**, 673-680.
- Seliger, H. H., Lall, A. B., Lloyd, J. E. and Biggley, W. H. (1982b). The colors of firefly bioluminescence—II. Experimental evidence for the optimization model. *Photochem. Photobiol.*, **36**, 681-688.
- Waterman, M. L., Adler, S., Nelson, C., Greene, G. L., Evans, R. M. and Rosenfeld, M. G. (1988). A single domain of the estrogen receptor confers deoxyribonucleic acid binding and transcriptional activation of the rat prolactin gene. *Molec. Endo.*, **2**, 14-21.
- White, E. H., Rapaport, E., Seliger, H. H. and Hopkins, T. A. (1971). The chemi- and bioluminescence of firefly luciferin: an efficient chemical production of electronically excited states. *Bioorg. Chem.*, **1**, 92-122.
- Wood, K. V., Lam, Y. A. and McElroy, W. D. (manuscript in preparation). Sequence determinants of color in beetle bioluminescence.
- Wood, K. V., Lam, Y. A. and McElroy, W. D. (1989a). Introduction to beetle luciferases and their applications. *J. Biolum. Chemilum.*, **4**, 289-301.
- Wood, K. V., Lam, Y. A., Seliger, H. H. and McElroy, W. D. (1989b). Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science*, **244**, 700-702.
- van Zonneveld, A. J., Curriden, S. A. and Loskutoff, D. J. (1988). Type 1 plasminogen activator inhibitor gene: functional analysis and glucocorticoid regulation of its promoter. *Proc. Natl. Acad. Sci. USA*, **85**, 5525-5529.

# Complementary DNA Coding Click Beetle Luciferases Can Elicit Bioluminescence of Different Colors

KEITH V. WOOD,\* Y. AMY LAM, HOWARD H. SELIGER, WILLIAM D. McELROY

Eleven complementary DNA (cDNA) clones were generated from messenger RNA isolated from abdominal light organs of the bioluminescent click beetle, *Pyrophorus plagiophthalmus*. When expressed in *Escherichia coli*, these clones can elicit bioluminescence that is readily visible. The clones code for luciferases of four types, distinguished by the colors of bioluminescence they catalyze: green (546 nanometers), yellow-green (560 nanometers), yellow (578 nanometers), and orange (593 nanometers). The amino acid sequences of the different luciferases are 95 to 99 percent identical with each other, but are only 48 percent identical with the sequence of firefly luciferase (*Photinus pyralis*). Because of the different colors, these clones may be useful in experiments in which multiple reporter genes are needed.

NEARLY ALL OUR KNOWLEDGE OF beetle luciferases is derived from studies of a single species, the North American firefly *Photinus pyralis*. Comparative studies with other beetle luciferases have been hampered because of limited availability of the other species. Evolutionarily, beetle luciferases are unrelated to any of the other groups of luciferases that have been studied biochemically (1). Little is known about the luciferases from other beetles except that they all catalyze the production of various colors of light through the oxidative decarboxylation of beetle luciferin (2). Since the substrates of the luminescent reaction are the same in all these beetles, the different colors must be due to differences in the structure of the enzymes (3).

Recently we cloned a cDNA that codes for the luciferase of *P. pyralis*, and have shown that it can be used to express bioluminescence in *Escherichia coli*. We report here the cloning of cDNAs that code for several new luciferases from a bioluminescent click beetle, *Pyrophorus plagiophthalmus*. This beetle is unusual because it can emit bioluminescence of a wide range of colors from a single species. The expression products in *E. coli* of the cDNAs derived from this beetle are able to produce green, yellow-green, yellow, and orange light. As determined from the nucleotide sequences of the clones, the amino acid sequences of these click beetle luciferases are highly conserved among one another, but diverge from the sequence of the firefly luciferase. Taxonomy indicates that the click beetle luciferases probably are the most evolutionarily distant of the beetle luciferases from the firefly

luciferase (4). This distance is reflected by differences in their chemical properties.

*Pyrophorus plagiophthalmus* is a large beetle with two sets of light organs. One set, on the dorsal surface of the head, emits light that is greenish but the exact color varies between individual beetles of the species, ranging from green (548 nm) to yellow-green (565 nm). The other set, at the anterior of the abdomen, generally emits light of a longer wavelength than the head organs but also varies between individuals ranging from green (547 nm) to orange (594 nm) (5). We converted mRNA isolated from the abdominal light organ of 60 beetles to cDNA and inserted this into a specialized lambda cloning vector, Lambda ZAP (6). The ability to convert this modified lambda vector into a bacterial expression plasmid (Bluescript) through an in vivo process allowed us to screen the cDNA library by two methods (7). In the phage form of the library, we screened with antibody to firefly luciferase that cross-reacts with the click beetle luciferases (8) and isolated four full-length clones that expressed bioluminescence in *E. coli*. A portion of the cDNA library was converted into the plasmid form, and we screened this for bioluminescence in the bacterial colonies. Bioluminescence can be initiated in colonies of *E. coli* expressing luciferase by adding luciferin to the media (9). Seven more cDNA clones were isolated by this method. It was determined visually that of the eleven clones, one produced green light, one produced yellow-green light, six produced yellow light, and three produced orange light.

Immunoblot analysis confirmed the production of full-length click beetle luciferase in *E. coli*. Despite some of these clones being detected with antibody to firefly luciferase during the library screening of plaques, we could not detect the gene products in blots made directly with *E. coli* lysates. The

expression of bioluminescence was improved by transferring the cDNA clones into a plasmid vector incorporating the *tac* promoter (10). A lysate from *E. coli* expressing the green-emitting luciferase from this vector was partially purified. After gel electrophoresis and blotting, a single antigenic band was revealed that comigrated with the native click beetle luciferase. Subsequently one cDNA clone from each of the four color-emitting groups was sequenced. An open reading frame was revealed in each that could potentially code a protein, the sequence of which correlated with the entire length of the sequence for firefly luciferase. Thus the complete protein coding regions of the click beetle luciferases were apparently contained within their cDNA clones.

Expression of bioluminescence from the *tac* vector yielded sufficient intensity, upon addition of luciferin to the media, to allow measurement of the spectral distribution from intact cells (Fig. 1). This confirmed the visual assignment of the 11 cDNA clones into four color groups. For each of the four colors, the bioluminescence spectrum is a single peak qualitatively similar to the spectra of the native click beetle luciferases (3). The range of colors from the clones is representative of the full range measured from the abdominal light organs of living beetles. However, there are colors emitted by the beetles, within the extremes of this range, that do not correspond to any of the clones (5). Thus other luciferase genes may not have been isolated. Spectra of the luciferases were also measured from partially purified preparations obtained from lysates of the *E. coli* expressing the cDNA clones.



Fig. 1. Bioluminescence from colonies of *E. coli* expressing the click beetle luciferases. Four streaks of *E. coli*, each consisting of hundreds of colonies, show the four colors of bioluminescence emitted by the different luciferases. The colonies were grown on nitrocellulose filters layered on top of nutrient agar. To initiate the bioluminescent reaction, the filters were removed from the agar and soaked with 1 mM luciferin in 100 mM sodium citrate, pH 5.0. The photograph was produced from a 2-s contact exposure of the colonies onto Ektachrome 64.

K. V. Wood, Y. A. Lam, W. D. McElroy, Department of Chemistry, M-001, University of California, San Diego, La Jolla, CA 92093.  
H. H. Seliger, Department of Biology, Johns Hopkins University, Baltimore, MD 21218.

\*To whom correspondence should be addressed.

Between pH 6 to 7, the spectra of these preparations were indistinguishable from those of intact cells. At pH 8 there was a slight broadening of the spectra for the green- and yellow-emitting luciferases. The firefly luciferase shows a large spectral shift between pH 6 to 8. At pH 8 its spectral maximum is at 560 nm, which shifts to 615 nm (red) at pH 6 with a decrease in the quantum yield (11).

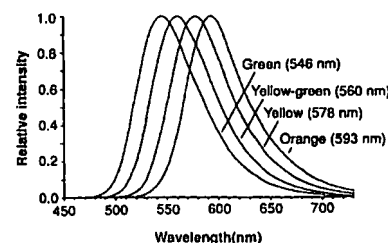
The sequences of the different click beetle luciferases are highly similar (Fig. 2). The open reading frame of each of the sequenced cDNA clones potentially codes a 543-residue polypeptide. Comparisons of the derived amino acid sequences show a 95 to 99% identity between the different color-emitting luciferases. Thus the number of amino acids that are responsible for the differences in the color is small. Because variation in color results directly from differences in the primary structures of the luciferases, specialized posttranslational modifications or unusual microenvironmental effects are not necessary to account for the color variation in the living beetles.

Comparison of the sequences of click beetle luciferases with that of firefly luciferase shows a low similarity. Alignment of their deduced amino acid sequences reveals that the various click beetle and the firefly luciferases are 48% identical (Fig. 3). Six gaps in the alignment of one to two amino acids in length account for most of a seven-amino acid difference in the lengths of the open reading frames between the firefly and click beetle luciferases. No regions in the alignment show especially high sequence similarity, thus giving little indication that particular regions have been conserved because of catalytic or structural constraints on the enzymes. An exception to this is in the last three amino acids which, for the firefly luciferase, have been shown to be necessary for translocation into peroxisomes (12). Given the close functional similarity of these enzymes, it is almost certain that the click beetle luciferases are also located in peroxisomes.

Firefly luciferase has historically been used as a bioluminescent reporter of chemical events associated with adenosine triphosphate (ATP) metabolism (13). With the cloning of its cDNA, this luciferase has also recently found application as an effective reporter of genetic events (14, 15). Its principal advantages are that (i) the initial polypeptide derived from the mRNA requires no posttranslational modifications for enzymatic activity; (ii) the luminescent reaction can be measured with high sensitivity; (iii) the assay of the gene product is rapid and does not use substrates requiring special precautions (such as radioactive isotopes or

chemically unstable compounds); and (iv) gene expression may be detected without disruption of living tissue. Compared with the conventionally used assay of chloramphenicol acetyltransferase (CAT) for gene activity, firefly luciferase is assayed in minutes as opposed to hours, and is 100 to 1000 times more sensitive (15).

The cDNAs coding for the click beetle luciferases also have these features, and, as they can be distinguished by color, may be



lysates of these *E. coli* between pH 6 to 8. The lysates were prepared from cells grown to middle or late log-phase growth at 37°C, and then for 2 hours at 30°C with isopropylthiogalactoside (IPTG) added to 1 mM. The cells were washed and resuspended in approximately 1/150 volume of the culture with 100 mM potassium phosphate, pH 7.0, 2 mM EDTA, and 35% of saturation  $(\text{NH}_4)_2\text{SO}_4$ . After lysis by sonication and removal of the debris by centrifugation,  $(\text{NH}_4)_2\text{SO}_4$  was added to 53% of saturation and the precipitate was dissolved in 1/15 volume of the lysate with 100 mM potassium phosphate, pH 7.0, 2 mM EDTA, and 50% glycerol. The spectra were measured from 10  $\mu\text{l}$  of this solution diluted 100-fold with 50 mM 2-(N-morpholino)ethanesulfonic acid, 50 mM 2-(N-morpholino)propanesulfonic acid (MOPS), 50 mM tricine, 5 mM  $\text{MgSO}_4$ , 1 mM EDTA, 0.1 mM ATP, 1 mM NaF, 0.2 mg of bovine serum albumin per milliliter, and 10% glycerol. The spectra were measured at pH 6.0, 7.0, and 8.0.

FF	EDAAIKKAAFY...G...QHKMKRYALVGTETFTAHIVNTATAYMSVREANKRYLNTMHRIVVSLQ...M	81
GR	MKREKNNVYGPFLPDLTAGEMLFRLKSHLPQ--ALVDVGEWISYKEFFETTCLLAQSLHMCYKMSDVVSCAENNRFFLP	90
YG	.....I.....P.D.SL.....A.....N.....I.....	90
YE	.....I.....K.....I.....P.D.SL.....A.....N.....	90
OR	.....I.....K.....I.....P.D.SL.....A.....N.....	90
FF	VILGLFVA...ADI...NER...LNS...QTV...VSKGQITNKKLPLOK...MSTKDYQPGNYT...VT...HLP...P...EYD...V	183
GR	ITAAHYICMIVAPNVEYIPDELCKVMSISRPQLVCTKMLKRVLEVQSRDPIKRIIILDAVENTHGCESLPFI--GRYDCNNAIFP	180
YG	.....S.....K.....I.....T.....	180
YE	.....S.....K.....I.....T.....	180
OR	.....S.....K.....I.....T.....	180
FF	ESF.RDKTI.L.MN...S...ALP...TA...FS...R...IP...N...I...DTAI...SVV...H.G...MFTT...LIC...F...VLMY...EE...L...R	275
GR	LAVDPVEQVAALICSSCTGCPKGVQTHRVNVCRLIALDPRVCTQLIPGVTVLVVLPFFHAFPGFSLNLCYFNVGLRVLNLRPDQEAFLK	272
YG	.....Q.....I.....A.....	272
YE	.....Q.....I.....EA.....V.....E.....E.....	272
OR	.....Q.....I.....EA.....V.....G.....E.....	272
FF	SL...KIQALLTLFSFA...T...I...N.H.IAS...G...S...CAVA...PH...Q...Y...T...I...LITPPEGDDP...AV...K...V	367
GR	AIQDYEVRSVINVPATILFLKSPFLVDKYLSSLRLECCGAFLAKEVAETAVKRLNLPGLRCCPGCTESTSANTHSLRDFKSGSLGRVTP	364
YG	.....IV.....V.....G.....	364
YE	.....IV.....V.....G.....	364
OR	.....IV.....V.....G.....	364
FF	FPF...VVLD...T...V...R...VR...IMS...P...NAL...K...IA...W...FI...LS...Y...S...	459
GR	LMAAKIADRETCKALGPNQVGLCIRGPVSKGVVNVVATKEAIDDDGNLHSGDPCYDDEHFPVVDYKLYKYGSGVAPAELEILL	456
YG	.....V.....	456
YE	.....K.....	456
OR	.....K.....	456
FF	QH.N.F.AG.A.L.DD...A.V...LEH...TM.E...IV...V.SQ.TTA.K...V...EV.KGL...LDA...IR...I...AK.GGK...	550
GR	KNPCTIRDAVVGIPDLSEAGELPSAFVVIQPKRITAKEVYDYLAERVSHTKYLGGVRFVDSIPRNVTKYI--TRK--E--LLKQLLEKSSKL	543
YG	.....K.....	543
YE	.....K.....	543
OR	.....K.....	543

Fig. 3. Alignment of the amino acid sequences of the click beetle and the firefly luciferases is shown to emphasize sequence differences. The sequence information is derived from the open reading frames of the corresponding cDNA clones. The identity of each luciferase sequence is indicated at the right of each line by a two letter code: FF, firefly; GR, green-emitting click beetle; YG, yellow-green-emitting click beetle; YE, yellow-emitting click beetle; and OR, orange-emitting click beetle. Only the sequence for the green-emitting click beetle luciferase is shown in entirety. Gaps in the alignment of this sequence are indicated by hyphens. Other luciferase sequences have letter designations only at sites where they differ from the green-emitting luciferase; where the sequences are the same there is a period. Numbers on the right indicate the position of the amino acid at the end of each line. Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

chemically unstable compounds); and (iv) gene expression may be detected without disruption of living tissue. Compared with the conventionally used assay of chloramphenicol acetyltransferase (CAT) for gene activity, firefly luciferase is assayed in minutes as opposed to hours, and is 100 to 1000 times more sensitive (15).

The cDNAs coding for the click beetle luciferases also have these features, and, as they can be distinguished by color, may be

useful in situations where multiple reporters are desirable. Expression in exogenous hosts should differ little between these luciferases because of their sequence similarity. Also, since the colors do not shift near physiological pH, the different luciferases can be distinguished in vivo as well as in vitro. Thus the click beetle luciferases may provide a dual reporter system that can allow two different promoters to be monitored within a single host, or for different populations of

cells to be observed simultaneously. The ability to distinguish each of the luciferases in a mixture, however, is limited by the width of their emissions spectra. From calculations based solely on the overlap of the spectra of the green- and orange-emitting luciferases, one luciferase in a mixture should be detectable in the presence of a 25-fold excess of the other.

library screening and D. R. Helinski for supplying pAD9 and his comments on the manuscript. Supported by NSF grant DMB 8603776 and Office of Naval Research grant N00014-86-K-0267. This

work is dedicated to the memory of Marlene DeLuca, a leader in the field of bioluminescence.

6 December 1988; accepted 7 March 1989

#### REFERENCES AND NOTES

1. J. W. Hastings, *J. Mol. Evol.* **19**, 309 (1983).
2. H. H. Seliger, I. B. Buck, W. G. Fastie, W. D. McElroy, *J. Gen. Physiol.* **48**, 95 (1964).
3. H. H. Seliger and W. D. McElroy, *Proc. Natl. Acad. Sci. U.S.A.* **52**, 75 (1964).
4. R. A. Crowson, *The Biology of the Coleoptera* (Academic Press, London, 1981).
5. W. H. Biggley, J. E. Lloyd, H. H. Seliger, *J. Gen. Physiol.* **50**, 1681 (1967).
6. *Pyrophorus plagiophthalmus* were collected in Jamaica and frozen in liquid N<sub>2</sub>. Polyadenylated RNA was isolated from the abdominal organ; 1 µg was converted to cDNA and inserted into Lambda ZAP to yield 700,000 recombinants, 5.5% of which could express luciferase antigens in *E. coli*. Of 18 phages chosen from the unamplified library by their reactivity with antibody to firefly luciferase, 4 could express bioluminescence in *E. coli*. The library was screened directly for bioluminescent activity from five portions that were amplified and converted into the Bluescript plasmid. Several bioluminescent colonies from each portion were identified by their ability to darken x-ray film; seven were determined as arising from independent cDNA clones. From two of the portions, two sets of colonies were judged as arising from independent clones based on widely different luminescent intensities, which were subsequently confirmed by restriction mapping.
7. J. M. Short, J. M. Fernandez, J. A. Sorge, W. D. Huse, *Nucleic Acids Res.* **16**, 7583 (1988).
8. G. Wienhausen and M. DeLuca, *Photochem. Photobiol.* **42**, 609 (1985).
9. K. V. Wood and M. DeLuca, *Anal. Biochem.* **161**, 51 (1987).
10. The plasmid pKW9 contains the *trp-lac* hybrid promoter (*lac*) and ribosome binding site (RBS) derived from pDR540 (Pharmacia LKB Biotechnology), and a transcriptional terminator derived from pAD9. The replicon portion of the vector is derived from pUC19. The cDNA clones were excised from Bluescript with the Bam HI and Xho I sites of the vector, and inserted into these sites in pKW9 between the *lac* promoter/RBS and the transcriptional terminator. Although the Bam HI site in pKW9 is immediately adjacent to the RBS, the 5' untranslated regions of the cDNAs and a portion of the polylinker region of Bluescript removed the translational initiation codon from the RBS by at least 50 bp. The first ATG encountered downstream of the *lac* promoter is the proper initiation codon for each of the cDNAs, thus reducing spurious initiations and the production of fusion proteins.
11. W. D. McElroy and H. H. Seliger, *Arch. Biochem. Biophys.* **88**, 136 (1960).
12. S. J. Gould, G. A. Keller, S. Subramani, *J. Cell Biol.* **105**, 2923 (1987); *ibid.* **107**, 897 (1988); S. J. Gould *et al.*, *J. Cell Biol.*, in press.
13. M. DeLuca and W. D. McElroy, Eds., *Methods in Enzymology* (Academic Press, Orlando, FL, 1978), vol. 57.
14. J. R. de Wet, K. V. Wood, D. R. Helinski, M. DeLuca, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7870 (1985); D. W. Ow *et al.*, *Science* **234**, 856 (1986); H. T. Tatum, T. Masuda, E. Nakano, *Agric. Biol. Chem.* **52**, 1123 (1988); P. K. Howard, K. G. Ahern, R. A. Firtel, *Nucleic Acids Res.* **16**, 2613 (1988); A. G. DiLella *et al.*, *ibid.*, p. 4159.
15. J. R. de Wet, K. V. Wood, M. DeLuca, D. R. Helinski, S. Subramani, *Mol. Cell. Biol.* **7**, 725 (1987); S. J. Gould and S. Subramani, *Anal. Biochem.* **175**, 5 (1988).
16. We thank S. T. Swanson for assistance with the

## Reexamination of the Three-Dimensional Structure of the Small Subunit of RuBisCo from Higher Plants

STEFAN KNIGHT, INGER ANDERSSON, CARL-IVAR BRÄNDÉN

The structure of L<sub>8</sub>S<sub>8</sub> RuBisCo (where L is the large subunit and S is the small subunit) from spinach has been determined to a resolution of 2.8 Å by using fourfold averaging of an isomorphous electron density map based on three heavy-atom derivatives. The structure of the S subunit is different from that previously reported for the tobacco S subunit in spite of 75 percent sequence identity. The elements of secondary structure, four antiparallel β strands and two α helices, are the same, but the topology and direction of the polypeptide chain through these elements differ completely. One of these models is clearly wrong. The spinach model has hydrophobic residues in the core between the α helices and β sheet as well as conserved residues in the subunit interactions. The deletion of residues 49 to 62 that is present in the *Anabaena* sequence removes a loop region in the spinach model. The positions of three mercury atoms in the heavy-atom derivatives agree with the assignment of side chains in the spinach structure.

CHAPMAN *et al.* (1) HAVE RECENTLY described the tertiary structure of plant RuBisCo, the key enzyme (2) in the Calvin cycle of carbon dioxide fixation in photosynthesis. Their model is based on an electron density map to 2.6 Å of the L<sub>8</sub>S<sub>8</sub> molecule from tobacco. We have determined the structure of L<sub>8</sub>S<sub>8</sub> RuBisCo from spinach to 2.8 Å resolution and find very significant differences in the structure of the S subunit compared with the reported tobacco structure. Since there is 75% identity between the amino acid sequences of these two polypeptide chains, they are expected to have similar tertiary structures.

Crystals of spinach RuBisCo that diffract to 1.7 Å resolution were grown from solutions of the activated form of the enzyme with a bound transition-state analogue (3). These crystals contain one-half the L<sub>8</sub>S<sub>8</sub> molecule in the asymmetric unit. There is a local noncrystallographic fourfold axis through the molecule, which has approximate 422 symmetry. X-ray data were collected on the synchrotron radiation source in Daresbury, United Kingdom, for the native enzyme and three heavy-atom derivatives. An initial electron density map was calculated with the use of isomorphous phase angles. These were refined by real-space averaging (4) around the local fourfold axis. Data collection procedures and phasing statistics have been briefly described (5).

The final electron density map was of very good quality, as would be expected by fourfold averaging of an electron density map

based on three heavy-atom derivatives. Almost all of the side chains could easily be identified from the known sequences of the spinach S and L chains (6, 7), which comprise 123 and 475 residues, respectively. The sequence of the S subunit, which was determined by amino acid analysis (6), contains only one Cys residue, Cys 112. However, two independent determinations of the amino acid content of the spinach small subunit (8) made in different laboratories have shown that there are three Cys residues per subunit. Furthermore, almost all of the small subunits from higher plant RuBisCo for which the sequences are known contain three Cys residues at positions 41, 77, and 112. We therefore conclude that in all probability the spinach small subunit also contains Cys residues at these three positions. Our electron density map also strongly supports Cys side chains at these positions; the side-chain electron densities are appropriate for Cys (Fig. 1b).

We first built the L chain (5) using the known structure of L<sub>2</sub> RuBisCo from *Rhodospirillum rubrum* (9). We found, in agreement with the work on the tobacco enzyme (1, 10), that higher plant L chains have a structure that is quite similar to that of the bacterial enzyme (9) except at the carboxyl terminal. The arrangement of the L subunits in the spinach enzyme into four L<sub>2</sub> dimers

Swedish University of Agricultural Sciences, Uppsala Biomedical Center, Department of Molecular Biology, S-751 24 Uppsala, Sweden.

## Cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica*

Li Ye<sup>1</sup>, Lys M. Buck<sup>2</sup>, Holly J. Schaeffer<sup>3</sup>, Franklin R. Leach<sup>\*</sup>

Department of Biochemistry and Molecular Biology, Oklahoma State University, 246 NRC, Stillwater, OK 74078-3035, USA

Received 22 May 1996; revised 4 November 1996; accepted 5 November 1996

### Abstract

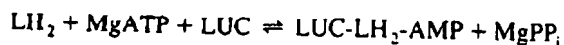
The first cDNA from the Photurinae subfamily of the Lampyridae encoding a firefly luciferase from lantern mRNA of *Photuris pennsylvanica* has been cloned, sequenced, the amino-acid sequence predicted and the sequence reported to GenBank. The cDNA was about 1.8 kb in length with the largest open reading frame coding for a 545-residue protein. The 5' noncoding region is 61 bp long and the 3' noncoding region is 135 bp in length. There is a 24-nucleotide poly(A) tail. When the amino-acid residues are aligned, *P. pennsylvanica* contains 154 (about 28% of the total residues) that are conserved in all 16 of the deduced luciferase sequences that are presently available. In this *P. pennsylvanica* luciferase, the amino acids at 276 of the positions are the same at corresponding positions of at least one of the other enzymes. There are two amino-acid differences between this luciferase and the unpublished sequence obtained by Dr. Keith Wood for a putative larval *Photuris* firefly luciferase cloned from a Maryland firefly. Signature amino-acid sequences and domains found in the deduced sequence are for adenylate kinase, the putative AMP-binding domain, luciferin 4-monooxygenase, 4-coumarate CoA ligase, long-chain fatty acid CoA ligase, 2-acylglycerophosphoethanolamine acyltransferase, the microbody-directing sequence, peptide-synthesizing complexes, and acyladenylate-synthesizing enzymes.

**Keywords:** Luciferase; Bioluminescence; Luciferin; Primary structure; (*P. pennsylvanica*); (Firefly)

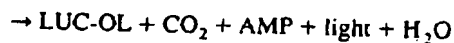
### 1. Introduction

Firefly luciferase (*Photinus* luciferin:oxygen 4-oxidoreductase, EC 1.13.12.7, abbreviated LUC) produces light by the oxidative decarboxylation of luciferin (LH<sub>2</sub>) as shown in the following equations that represent the two-step reaction.

Step one forms an enzyme-bound luciferyl adeny-



Step two is the oxidative decarboxylation of luciferin with the production of light upon decay of the excited form of oxyluciferin



There is a slow release of the oxyluciferin product, OL, from the enzyme-product complex.

The enzymatic reaction has a quantum yield of 0.88 photon/molecule of luciferin oxidized [1]. The

<sup>\*</sup> Corresponding author. Fax: +1 (405) 7447799; E-mail: firefly@biochem.okstate.edu

<sup>1</sup> Present address: Department of Molecular Biology, Holland Laboratory, American Red Cross, Rockville, MD, 20855.

<sup>2</sup> Present address: 1525 SW 7, #8, Albany, OR 97321.

<sup>3</sup> Present address: Department of Horticulture, Michigan State University, East Lansing, MI 48824.

Table 1  
Cloned firefly luciferases

Year	Species	Abbrev. LUC	GenBank sp	PIR	Locus	% same	# Amino acids	A <sub>max</sub>	Reference
1987	<i>Photinus pyralis</i>	Ppy	M15077	A26772	PPYLUC	60.4	550	562	[7]
1989	<i>Pyrophorus plagiophthalmus</i> (4)								
	Greco-LucGR	PplGR		S29352				546	[8]
	Yellow green-LucYG	PplYG		S29353		48.0	543	560	
	Yellow-LucYE	PplYE		S29354		48.9	543	578	
	Orange-LucOR	PplOR		S29355		49.3	543	593	
1989	<i>Luciola cruciata</i>	Lcr	M26194	S0181	FELLUC	49.1	542	562	[9]
1992	<i>Luciola lateralis</i>	Lla	X66919	S23437	LLLUCI	55.3	548	552	[10]
1993	<i>Luciola mingrelica</i>	Lmi	S61961	S33788	S61961	55.5	548	570	[11]
1995	<i>Photuris parvula</i>	Hpa	L39929		HOTLUCI	53.7	548	568	[12]
1995	<i>Pyrocoelia miyako</i>	Pma	L39928		PIBLUCIF	55.3	548	550	[12]
1995	<i>Lampyris noctiluca</i>	Lno	X80479		LNLUCPROT	60.1	547	550	Gene bank, [13]
1995	<i>Luciola lateralis</i> gene	Lla(g)	Z49891		LLLUCIFM1	61.7	548	552	Gene bank
1994	<i>Photuris pennsylvanica</i>	Ppe(LY)	U31240		PPU31240	56.2	545		This lab, [14]
1994	<i>Photuris pennsylvanica</i>	Ppe119				58	552	560	Patent
	<i>Photuris pennsylvanica</i>	Ppe(KW)				55	552	538	Wood **
	<i>Photuris pennsylvanica</i>	Ppe2(KW)				100	545	538	Wood **

From GenBank, Swiss Protein Sequence Data Base at Expasy (sp), and Protein Identification Resource at Johns Hopkins (PIR).  
\* CAS-registry 160831-30-1 protein sequence for JP 94303982 A2, not entered into Chemical Abstracts until after Li Ye thesis was submitted [14]; \*\* Keith Wood, personal communication. The abbreviations for the LUC for specific organisms are adapted from those used by Wood (personal communication). Where both a cDNA and a gene are deposited, the gene sequences is indicated by a (g) after the standard abbreviation. When the same species has two or more sequences published, a Roman numeral is added. For sequences obtained in different laboratories the initial of the principal investigator are added in parentheses. When luciferase from one species produce more than one color light, a two letter abbreviation code suggested by Wood et al. [8] is used. The % same column shows the percentage of identity with *P. pennsylvanica* sequence. The wavelength of light produced is based on that by the beetle unless more than one species exist and then it is based on that determined using the enzyme.

enzyme is widely used in the quantitation of ATP and other biochemically important compounds [2–5] and as reporter of gene expression (see bibliographic lists in *J. Biolumin. Chemilumin.* 5, 141–152 (1990) and 8, 267–291 (1993)).

The gene for firefly luciferase of the North American firefly, *Photinus pyralis*, was cloned and sequenced by DeLuca and colleagues [6,7]. Since those studies, firefly luciferase cDNAs or genes have now been cloned from several other beetle species; these are listed in Table 1. This paper reports the cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica*; this is the first species from the Photurinae subfamily to have its luciferase sequence determined and reported to GenBank. *Photuris pennsylvanica* is a twilight/night-active firefly while the common, well-characterized North American species, *P. pyralis*, is dusk-active, flashing only during twilight. Before mating, *Photuris* fireflies females respond to courtship signals of conspecific males [15]. After mating, they become 'femmes fatales' by answering the courtship flashes of males of other species who are then treated as prey. Predation by aggressive mimicry is known only for *Photuris* [16].

The *Photuris* genus has been mainly studied from the biological standpoint. The biochemical and structural changes that occur during light organ development have been studied by Strause et al. [17]. During development the larval light organ regresses and is replaced by the adult lantern. During pupation the levels of luciferase and luciferin remain constant in the posterior half of the pupa while there is an initial increase followed by a decrease of luciferase and luciferin in the anterior half. Strause and DeLuca [18] found a luciferase isozyme in larval *Photuris pennsylvanica* that is distinct from the enzyme of the adult. This laboratory has identified two firefly luciferases from adult *P. pennsylvanica* lanterns during Sephadex G-150 chromatography [14].

Fig. 1. Nucleotide and predicted amino acids sequences for cDNA from *Photuris pennsylvanica*. The 61 bp leader is shown in groups of 10 bp with a residual. The coding sequence starts at bp 62 and continues through bp 1700. The 5' and 3' noncoding region is shown in groups of 10 bp. The figure was prepared by editing a DNA Strider report after the consensus sequence was derived using the AssemblyLIGN program.

```

GGGCGGCGGCT AGACTTTCCT GGCAGGCTCTG ATCTAAATAC TAAATATAC AGCTGAATAT A
1/1 31/31
ATG GGA GGT AAA AAC ATT CCA CAC GGA GTC GGA CCA TTC CAC CCG GTC GAT GGC ACT
H E D E H I L Y G D E P F H P L A D D T
41/31 91/31
GTC GGA GGA CAG AGC TCC CAC GCA CCA CCA CCG CAC GCA GTC ATT CCA GGA GTC ATT GCA
A G E Q H P Y A L S R Y A C I S G C I A
121/41 151/31
CTG ACA AAC GTC CAC ACA AAA GAA AAT GTC CCA CAC GAA TGA TTC CCA AAA CTG GCG CTT
L T M A N T K E N V L Y E E P L R L S C
181/41 211/71
TGT CCA GCG GAA AGT CTT AAA AAG TAT GGA CCA AAA CAA AAC GAC ACA ATA GCG GTC CTT
R L A E S P K K Y D L K D H O T I A V C
241/81 271/91
AGT GAA AAT GTC CTG CAA TTC TCC CTT CTT CCA ATT GCA CCA CTG CAC CTT GGA ATA ACC
S E N G L Q F F L P C C A S L Y L G T C
301/101 331/111
GCA GCA CTT GTC AGT GAT AAA AAG ATA GAA GAT GGA CCA ATA CAC AGT CTT GTC ATT GCA
A P V S D K Y I E R E L C H S L G C I V
361/121 391/131
AAA CCA CCG ACA ACC TCC CCG ACC AAC CCA AAC CCA AAA GCA CTG AAT GCA AAA CTT
K P R C C P C S K N T F O K V L N V R S
421/141 451/151
AAA CCA AAA CAC GCA GAA ATT ATT CCA CCA GAC CCA AAT GAA GAC CCA GAA GTC TAT
F L K V Y S T I L C L D C H S D L O G Y
481/161 511/171
CAG CTC CAC AAC AAC TCC ACC CCA AAC CCG GAT ATT AAT CTC GAC GCA AAG AAA TTC
Q C L N H F I S Q N S D H L D V K K F
541/181 571/191
AAA CCA AAC CTT AAC CCA GAC GAT CAG GTC GCG GTC GCA CTG TCC TCC CTT ACA
K P M E P N R D D Q V A L N P S L S C
601/201 631/211
ACT GTC CTC TCC AAG GGA GTC ATG CCA ACC CAC AAG AAT ATT GTC GCA CCA CTT CTT CAC
T Q V S K G V N L T H E N C V A R F S H
661/221 691/231
CTC AAA GAT CTT ACT CTT GTC AAC GCA ATT AAC CCA ACT CCA CCA ATT CCA GCA ATA
C R D P T P G N A I S P T A L T V I
721/241 751/251
CAG CTC CAC CAC GTC CTT GTC AAG ACT ACC CCA CCA GAA TAT TCC ACT CCG GTC CCA
P F H N C P S N T T T L G C F T C D F R
781/261 811/271
GTC CTC CCA AGC CAC AGC CTT GAA GAA AAA TTA CTC CCA CCA CCA CCA CCA CCA CCA
V A L N H T F E E K L P L C S L Q D Y K
841/281 871/291
GTC GAA AGT ACT TTA CTC CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA
V S S T L L V P T L H A F P K S L L V
901/301 931/311
GAA AAG CAC GTC CCA CTG CAC CCA AAA GAA ATT GCA CTT GTC GCA CTT CCA CCA AAA
E K Y D L S H L K E C A S C G A P L S K
961/321 991/331
GAA CCG GAG AGC GTC AAA AAA CCG CTT AAA CCA AAC TTT GTC AGT CCA GTC CCA GTC
R S G E H V K R R F E L N F V R Q G C
1021/341 1051/351
CTA ACA GAA ACT ACT CCG GTC CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA
L T E T S A V L C T P C D V R P G S
1081/361 1111/371
ACT GTC AAA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA
T G K I V D F H A V K V V D P T T G K I
1141/381 1171/391
CTG GCG CCA AAC GAA ATT GGA GAA CTG CAC CTT AAA GGC GAC AGC ATA CTG AAA AGT CAC
L G P H E T G E L Y F K G D H I K S Y
1201/401 1231/411
CAT AAC AAT GAA GAA GTC ACT AAA GCA ACC ACC AAC AAA GAC GGA CTG CTG CTT GTC
Y M H E E A T K A C C H K D G M L R S N
1261/421 1291/431
GAC ACT GTC CAC CAC CAC AAT GAT GGC CAC CTT CAC ACT GTC GAC AGT CAG CAG CCA
D I A Y T D N D G H F Y I V D R L K S L
1321/441 1351/451
ATC AAA TAT AAA GTC TAT CAG GTC CCA CTT GTC GAA ATT GAG GGC ACA CTT CCA CCA CAC
I E Y K G Y Q V A P A E I E O I L L Q H
1381/461 1411/471
CAG CAC AAT GTC GAC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
P Y I V D A G V T G I P D E A A S G L P
1441/481 1471/491
GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
A A D V V V D T G K Y L N E Q C V Q N F
1501/501 1531/511
GTC CCG AGT CAA GTC CCA ACA CCA AAA TGG CCA GTC GTC GTC GTC GTC GTC GTC GTC
V S S Q V S T A K M L R C G V K P L D S
1561/521 1591/531
ATC CCG AAA GGA CCA ACT GCA AAA ATT GAC GGA AAA GTC CCA GGA CAA AGT CTT GAA AAA
I P E G S T G K C D R K V L R D H P E K
1621/541
CAC AAA CTT AAG CTG CAA
H E S K L
GGGCGGCGGCT AGACTTTCCT GGCAGGCTCTG ATCTAAATAC TAAATATAC AGCTGAATAT A
CAGTGAATATAG CAGTGAATATAG CAGTGAATATAG CAGTGAATATAG CAGTGAATATAG CAGTGAATATAG

```

The overall goal of this study is to increase knowledge concerning amino acids that might function in catalyzing light production by luciferin oxidation by obtaining an unique sequence from another subfamily of fireflies. Comparison of the amino-acid sequences of several firefly luciferases may indicate which amino acids are functionally important since these should be conserved among the various species. The sequence of *P. pennsylvanica* luciferase has been deposited in the GenBank as entry U31240.

## 2. Materials and methods

### 2.1. Bacterial strains and media

*E. coli* strain XL1-Blue MRF' (Stratagene Cat. # 200301) was used for λZAP Uni-Zap™ library plating. *E. coli* strain XL1-Blue (Stratagene Cat. # 200268) was used for amplification of λZAP Uni-Zap™ library and for recombinant plasmid replication and preparation. *E. coli* strain SOLR™ (Stratagene Cat. # 200298) was used as a plasmid host for in vivo excision and cDNA library screening.

### 2.2. Firefly collection

Fireflies were collected locally with nets and stored in jars until the evening's collection was completed. The live fireflies were taken to the laboratory, sorted according to species, and immediately frozen in liquid nitrogen. The frozen fireflies were stored in a –80°C freezer.

### 2.3. mRNA isolation and library preparation

The lanterns of the frozen *P. pennsylvanica* fireflies were removed under liquid nitrogen and then one g of lanterns was pulverized under liquid nitrogen with a mortar and pestle. The lantern RNA was isolated by using a Stratagene RNA Isolation Kit

(Cat. # 200345, Lot 138) and then the mRNA was isolated using Stratagene's Quik™ mRNA Isolation Kit (Cat. # 200349, lot 133). A Stratagene ZAP-cDNA™ Synthesis Kit (Cat. # 200400, Lot UC122) was used for construction of the cDNA library using 5 µg mRNA. In vivo excision of pBluescript® phagemid from Uni-ZAP™ XR (Stratagene ZAP-cDNA™ Synthesis Kit) was done according to the instructions provided to produce a phagemid.

### 2.4. Expression

Expression was achieved by following a procedure similar to that of Devine et al. [11]. The instructions supplied with Stratagene's ExAssist™ kit were followed for excision. Twelve 150 mm plates were used for the first screening and incubated at 37°C overnight. The colonies on each individual plate were lifted onto a nitrocellulose filter and the filters were placed on fresh LB-ampicillin plates. The master plates and lifted filters were incubated at 37°C for another 2 h. The filters with colonies were switched to 22°C and incubated for another 2 h for expression. The filters were removed from the plates, soaked with 1 mM MgSO<sub>4</sub>, 2 mM EDTA, pH 5.0 for 5 min, wrapped with Saran Wrap® and exposed to X-ray film overnight. After film development, the positive colonies (bioluminescent) were identified (they also could be observed by eye after dark-adaptation). Two colonies were picked from the master plates, and streaked onto fresh LB-ampicillin plates. Positive colonies were purified by two more rounds of plating and screening.

### 2.5. Sequencing

The pBluescript® phagemid was used for sequencing. DNA sequencing was done by the Sarkeys Biotechnology Laboratory in this Department. The sequencing was started using T3 and T7 primers.

Fig. 2. Comparison of the amino-acid sequences from sixteen firefly luciferases. An alignment of the amino-acid residues from the sixteen firefly luciferase listed in Table 1 was prepared. The shading used indicate: ■ an amino-acid residue conserved in all 16 luciferases; □ a residue found in *Photuris pennsylvanica* and at least one other firefly luciferase sequence; and –, a gap in the sequence used for alignment of the residues. The alignment was handmade and checked by comparison to an alignment made using the MACAW, and Clustal programs.







Subsequent sequencing was continued in both directions using primers designed using the Oligo<sup>®</sup> 4.06 program (National Biosciences, Inc.) based on the determined sequence. The primers were synthesized by the Sarkeys Biotechnology Laboratory.

## 2.6. Measurement of luciferase gene size

The plasmid containing the luciferase gene was digested with *Eco*R I and *Xho*I at 37°C overnight. A separate digestion was done using *Bam*HI and *Kpn*I under the same conditions. The digested samples were separated by 1% agarose gel electrophoresis with a 1 kb DNA ladder (BRL, Cat. No. 1561057) as a standard and uncut plasmid as a control.

## 2.7. Computer analysis of data

The following programs and data bases (versions) were used: AssemblyLIGN, V 1.0.7; Beauty [19], BLASTPAT, and various other search algorithms via the BCM Search Launcher of the Human Genome Center at the Baylor College of Medicine, Houston TX; Blocks [20,21]; DNA Strider, V1.2 [22]; GenBank, NCBI, release 89.0 [23]; MacVector, V 4.5.2 [24]; pI/MW calculated at the ExPASy Server; ProDom, release 28 [25,26]; ProSite [27]; PROSITE, release 13 [27]; and SWISS-PROT, release 21.0 [28].

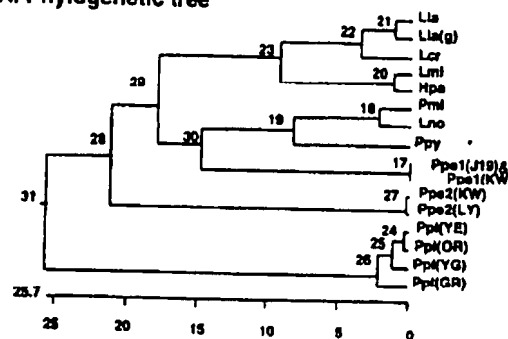
## 3. Results

### 3.1. Isolation and sequencing of cDNA

The cDNA library was prepared from the lanterns of locally collected *P. pennsylvanica* fireflies, expressed in *E. coli*, and screened for light production after luciferin addition. Since the screening detected expressed bioluminescence, only functional cDNA sequences were identified. The insert size as determined after *Eco*RI and *Xho*I digestion and 1% agarose gel electrophoresis was about 1.8 kb which is sufficient to code for the entire luciferase polypeptide of approx. 550 amino-acid residues (based on the length of other firefly luciferases). In a restriction enzyme-based analysis, the selected clone did not contain a *Kpn*I restriction site but did have a single *Bam*HI site within the luciferase-coding sequence

(which is not found in the *Photinus* sequence). No *Kpn*I restriction site is present in the *Photinus* sequence. Devine et al. [11] found clones of *Luciola mingrelica* cDNA that had either one or two *Kpn*I sites.

### A. Phylogenetic tree



### B. Biological classification

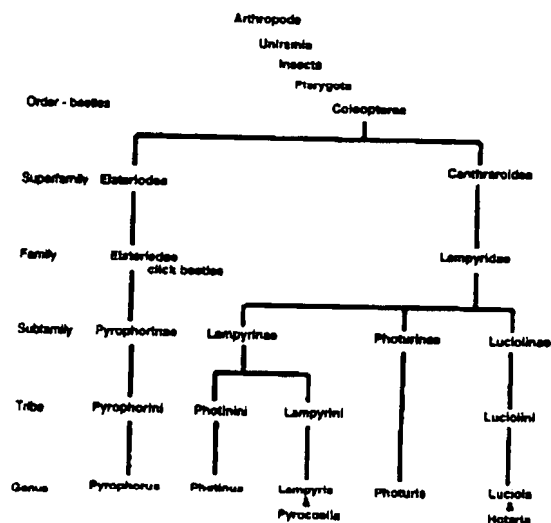


Fig. 3. Relationships among the amino-acid residues and the biological classification of the sixteen sequenced firefly luciferases. (A) The relationship among firefly species based on firefly luciferase amino acid sequences as determined using the protein parsimony and the Clustal algorithm in the DNA Star program. The length of each pair represents the distance between the sequence pairs and the scale beneath the tree measures the genetic distance between sequences. (B) Biological classification of the firefly species from which the luciferase has been sequenced. This classification of firefly species was adapted from Herring [29] and Campbell [5].

Table 2  
Amino-acid occurrences in sixteen firefly luciferases

	Firefly luciferase source															
	Ppy	Lcr	Lja	Ljag	Lmi	PpIGR	PpIYE	PpIOR	PpIYG	Hpa	Pmi	Lno	PpeI(J19)	PpeI(KW)	Ppe2(KW)	PpeI(Y)
MW	60706	59979	60087	60010	60456	60652	60412	60296	60268	60327	60917	60365	60931	60931	60610	60610
pI	6.42	7.07	6.50	7.96	6.24	6.69	6.39	6.69	6.69	6.27	6.12	6.09	7.55	7.55	8.39	8.40
Amino acid																
Total 550	548	548	548	548	548	543	543	542	542	548	548	547	552	552	545	545
Nonpolar																
A	42	32	34	34	33	36	37	37	37	33	36	39	36	36	34	33
C	4	8	7	7	8	13	13	13	13	8	9	9	11	11	7	7
F	30	23	23	23	27	23	24	24	24	26	29	28	25	25	30	30
I	38	33	36	36	36	39	42	43	38	35	39	39	42	42	39	39
L	52	49	49	49	48	57	54	54	56	49	49	49	54	54	54	54
M	14	11	12	12	13	11	11	11	11	13	18	17	9	9	11	11
P	29	29	28	29	30	28	28	28	28	30	26	28	30	30	22	23
V	44	55	49	49	50	49	47	45	48	50	39	37	41	41	38	38
sum	253	240	238	239	245	256	256	255	255	244	245	246	248	248	235	235
No charge																
G	45	51	52	53	49	38	38	39	39	50	46	49	46	46	43	43
N	19	20	20	20	19	21	23	23	23	19	13	14	21	21	25	25
Q	16	13	12	11	14	13	14	14	14	16	20	19	18	18	17	17
S	29	30	29	28	26	31	32	31	31	25	21	32	32	32	33	33
T	29	36	34	34	32	19	19	19	19	35	31	29	29	29	34	34
W	2	1	1	1	1	3	2	2	2	1	2	1	1	1	2	2
Y	19	21	22	22	19	21	20	20	20	20	21	22	24	24	24	23
sum	159	172	170	169	160	146	148	148	148	196	154	166	171	171	178	177
Charged																
D	31	25	27	28	25	26	25	25	26	24	32	30	23	23	27	27
E	33	39	40	37	42	39	40	39	38	41	33	34	36	36	31	31
H	14	8	8	8	15	13	13	13	13	13	17	14	14	14	12	13
K	40	43	44	46	42	35	37	38	37	42	36	37	42	42	47	47
R	20	21	21	21	19	28	24	24	25	18	31	20	18	18	15	15
sum	138	136	140	140	143	141	139	139	139	138	149	135	133	132	132	133

The amino-acid composition and molecular weights were calculated using the DNA Strider program with input data from GenBank, patent information, personal communications, and PIR. pI was calculated by the Expasy pI/MW program.

Table 3  
Percentage of identity and similarity among pairs of the 16 firefly luciferases

	PPy	Lmi	Lcr	Llu	Lldg	Ppl(GR)	Ppl(YG)	Ppl(YE)	Ppl(OR)	Pmi	Hpu	Lnu	Ppe(U19)	Ppe(KW)	Ppe2(KW)	Ppe2(LY)
PPy																
Lmi	67.9 (82.1)															
Lcr	80.8 (90.1)	60.4 (82.1)														
Llu	93.6 (96.9)	82.1 (90.5)	88.6 (96.9)													
Lldg	99.1 (99.5)	81.5 (90.3)	82.8 (90.3)	68.2 (82.8)												
Ppl(GR)		48.2 (67.2)	48.7 (67.5)	48.6 (67.6)	48.7 (67.7)											
Ppl(YG)		47.8 (67.7)	49.0 (67.8)	48.9 (67.1)	47.9 (67.3)	48.4 (67.7)										
Ppl(YE)		46.1 (98.5)	49.9 (67.8)	48.9 (68.1)	48.9 (67.3)	48.0 (67.8)	48.7 (67.7)									
Ppl(OR)		98.5 (99.5)	99.1 (99.5)	99.1 (99.5)	99.1 (99.5)	99.1 (99.5)	99.1 (99.5)	99.1 (99.5)								
Pmi																
Hpu																
Lnu																
Ppe(U19)																
Ppe(KW)																
Ppe2(KW)																
Ppe2(LY)																

Calculated using the GCG program BestFit. Shown in parentheses are the calculated similarities.

Fig. 1 shows the nucleotide sequence of the *P. pennsylvanica* cDNA and the deduced amino-acid sequence for the largest open reading frame. From the sequence analysis, the cDNA is 1831 bp long with an open reading frame (ORF) of 1635 bp. The ORF encodes a protein of 545 amino acids with a calculated molecular weight of 60 610 daltons. The 5' untranslated region contains 61 bp and the 3' untranslated region has 135 bp. The 3' noncoding region contains a poly(A) tail of 24 nucleotides.

### 3.2. Comparison of the deduced amino-acid sequences

The amino-acid sequences deduced from the 16 cDNAs and genes sequenced for firefly luciferases have been aligned (see Fig. 2) to allow determination of conserved amino-acid residues and suggest possible functional portions. There are 154 residues conserved among all the luciferases (about 28% of the total residues). In the putative *P. pennsylvanica* larval luciferase, the amino acids at 276 positions are the same at corresponding positions of at least one other species. One hundred and fifteen amino-acid residues are unique to the putative *P. pennsylvanica* larval enzymes. Of these, 24 residues are conserved in all other species.

The amino-acid compositions of the sequenced firefly luciferases are compared in Table 2. The calculated isoelectric points of putative *P. pennsylvanica* larval enzymes are the highest of all isozymes.

### 3.3. Relatedness

The amino-acid sequences of the firefly luciferases were analyzed by protein parsimony using the DNA Star program (Fig. 3A). The phylogenetic classification for these beetles is shown in Fig. 3B [5,29]. As expected, the various *Luciola* species are closely related. The relationships among the various species as determined by luciferase amino-acid sequence appear similar to the relationships based on biological classification. The percentage of identity and similarity were calculated for the 16 firefly luciferase sequences using the GCG program BestFit with a gap weight of 3.0 (Table 3). The *Ppe1* and *Ppe2* luciferases are 57% identical.

## 4. Discussion

### 4.1. Related sequences

When either the cDNA sequence or the predicted amino-acid sequence was used as the input sequence for computer-based searches for similarity, the high-scoring related sequences were the luciferases from both the Lampyridae and the Elateridae families, 4-coumarate CoA ligase, long-chain CoA ligase, 2-acylglycerophosphoethanolamine acyltransferase, and the peptide-antibiotic-synthesizing enzymes such as gramicidin S synthetase and tyrocidine synthetase. These relationships have been reported [30–38] for the other firefly luciferases.

### 4.2. Domain structure

A domain structure map for the predicted amino-acid sequence of *Photuris* firefly luciferase was developed by using the ProSite, ProDom, PRINTS, BLOCKS, and PepPepSearch programs. Fig. 4 illustrates these results and Table 4 defines the sites and their presumed functions, if known.

The T-250TLGYFT-256 sequence (see also Fig. 2) is the AMP binding block BL00455B. PS00339, the AA tRNA ligase II.2 sequence whose consensus sequence is [GSTALVF]-[DENQHRKP]-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY], was found as F-431YIVDRLKSL-440 (correct in 9/10 positions). The ProSite convention is {ambiguities where indicated amino acids are acceptable} and {amino acids not accepted in this position}. The G-338YGLTRYSAVLITPDTDVRPGSTG-362 sequence is the domain II that is conserved in acyl-adenylate-synthesizing enzymes [30]. The adenylate kinase signature, PS00113, with consensus sequence [LIVMFYW](3)-D-G-[FY]-P-R-x(3)-[NQ] was tentatively found as I-411NKDGWLRSGDI-422 (correct in 7/12 positions). The G-415WLRSGD-421 sequence is the domain III that is conserved in acyl-adenylate-synthesizing enzymes [30]. The SKL sequence at the C-terminus is the microbody-directing sequence and was detected by the SORT program (it is also ProSite PS00342).

The putative AMP-binding domain signature of [LIVMFV]-x(2)-[STG](2)-G-[ST]-[STE]-[SG]-x-[PALIVM]-K (ProSite P00455) was found as V-194MFSSGTTGVSK-205 and is highly conserved

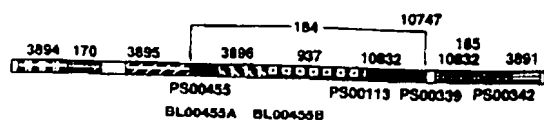


Fig. 4. Putative functional domains of *Photuris pennsylvanica* firefly luciferase. The figure is a composite of the BEAUTY map and ProDom. ProSite and Blocks maps of possible functional domains. The ProDom sequences are: 3894, luciferin 4-monooxygenase (firefly luciferase); 170, 4-coumarate CoA ligase; 3895, luciferin 4-monooxygenase; 184, gramicidin S synthetase; 3896, luciferin 4-monooxygenase; 937, 4-coumarate CoA ligase; 10747, 4-coumarate CoA ligase; 10748, 4-coumarate CoA ligase; 10832, 2-acylglycerophosphoethanolamine acyltransferase; 185, gramicidin S synthetase; and 3891, luciferin 4-monooxygenase. The ProSite sequences are: PS00455, putative AMP-binding domain sequence; PS00113, adenylate kinase signature (similar); PS00339, aminoacyl-tRNA synthetase class II.2 (similar); and PS00342, microbodies C-terminal targeting sequence. The Blocks are: BL00455A and BL00455B, the putative AMP-binding domain.

number in parentheses is the number of occurrences in the 16 sequences).

Among the sequences not found (whose possible existence was considered on the basis of similarity in function) is the consensus pattern of [FYH]-R-X-[DE]-X(4, 12)-[RH]-X(3)-F-X(3)-[DE], ProSite PS00179 which is the amino-acid tRNA ligase II 1 pattern. ProSites that were sought but not found include: the ATP/GTP-binding site motif A (P-loop) PS00017, the chloramphenicol acetyltransferase active-site PS0100, the protein kinase signatures (PS00107, PS00108, and PS00109), the ubiquitin-activating enzyme signatures (PS00536 and PS00865), and the acyl-CoA-binding protein signature PS00880. Since the P-loop occurs in many proteins that bind ATP/GTP, it was the subject of a search. Protein kinases have binding sites for ATP and there is some similarity. The CoA binding domains were sought because CoA influences the time-course of light production by firefly luciferase [39–41].

The domains that can be recognized involve ATP(AMP)-binding sites, regions that interact with ATP, regions that are involved in reactions leading to

among the various firefly species — the consensus sequence for luciferase is  $I_{(11)}M_{(12)}N_{(10)}S_{(16)}S_{(16)}G_{(16)}S_{(10)}T_{(16)}G_{(16)}L_{(14)}P_{(14)}K_{(16)}$  (where the subscript

Table 4  
ProDoms and ProSites found in firefly luciferase

Element	Type	Length, residue	Number proteins	% identity with <i>Ppe2</i>	Identified function
<b>ProDom</b>					
3891	Firefly luciferase	30	3	60	Unknown
3894	Firefly luciferase	50	3	52	Unknown
3895	Firefly luciferase	61	3	49	Unknown
3896	Firefly luciferase	51	3	67	Unknown
184	Gramicidin S synthetase	34	45	50	Unknown
185	Gramicidin S synthetase	43	45	42	Unknown
937	4-coumarate-CoA ligase	25	13	45	Unknown
10748	4-coumarate-CoA ligase	27	1	52	Unknown
10747	4-coumarate CoA ligase	27	1	44	Unknown
10832	2-acylglycerophosphoethanolamine acyltransferase	24	1	42	Unknown
<b>ProSite</b>					
PS00113	Adenylate kinase	12	45	58	Adenylate kinase signature
PS00339	AA tRNA ligase II	10	99	80	Signature
PS00342	Firefly luciferase	3	88	100	C-terminal microbody-directing
PS00455			85		AMP-binding signature
motif 1		12		69	
motif 2		9		67	

% identity with *Ppe2* is based on the ProDom search or the listed amino-acid residues for the ProSite. ProDom 170 (gramicidin S synthetase) shows two firefly luciferases listed as members of the family, but there appears to be insufficient sequence homology over the whole sequence.

the formation of adenylate intermediates, and function in peptide synthetases. As expected, there are several regions that are found in other firefly luciferases.

#### 4.3. Evolutionary relationships

Wood [42] has reviewed the information on chemical mechanism and evolutionary development of the beetle luciferases. The level of dissimilarity among the beetle luciferases is large; only 27% of the amino-acid sequence is conserved among the cloned beetle luciferases. Wood concluded that the rate of evolution of the luciferases is high relative to other enzymes.

Wood presented a tree diagram showing the relationship of the beetle luciferases with the CoA synthetases [42]. The closest enzyme to the luciferases is 4-coumarate:CoA ligase which has 17% identity in amino-acid sequences. Wood [42] postulates that the firefly luciferases may have evolved from the CoA synthetases. CoA influences the pattern of light production by the luciferases. Without CoA, saturating concentrations of ATP produce a flash of light. CoA prevents the subsequent inhibition of light production and allows a sustained production of light. Wood [40] and Ford et al. [41] found that the -SH group of CoA was required for stimulation of activity. Other nucleotide analogs can produce a steady light production similar to that obtained with CoA [43]. These results have been interpreted as an enhanced turnover of the enzyme mediated by conformational changes. These observations suggest that there may be a vestigial CoA binding site on the luciferases. The crystal structure of Type III chloramphenicol acetyltransferase has been solved at a 1.75 Å resolution by Leslie [44] and the acetyl CoA binding site analyzed [45]. The amino-acid residues interacting with CoA are F-55, Y-56, F-104, S-148, W-152, K-177, Y-178, H-195, and D-199. A similar pattern of residues occurs in *P. pennsylvanica* and other firefly luciferases (the residue number corresponds to the *P. pennsylvanica* sequence and the ratio is the number of times that the residue is found in the aligned position in the 16 luciferases of known sequence): F330, 12/16; Y339, 12/16; F367 6/16 and at position 368, 8/16; W416, 16/16; S419, 11/16 or T, 5/16; K442, 16/16; Y443, 16/16; H460, 12/16;

and D465 16/16. These amino acids may constitute the CoA-binding site.

#### 4.4. Crystal structure

The crystal structure of recombinant *P. pyralis* luciferase has been determined at a resolution of 2.0 Å [46]. There are two compact domains — the N-terminal 80% and the C-terminal 20%. The N-terminal domain contains a  $\beta$ -barrel and two  $\beta$ -sheets which are flanked by  $\alpha$ -helices. The C-terminal region contains two short antiparallel  $\beta$ -strands and a three-stranded mixed  $\beta$ -sheet, with three helices packed against the side. There is a large cleft between the two parts (domains). Amino acids conserved in the firefly luciferases are in both parts. The P-loop is in a loop connecting antiparallel strands 6 and 7 of  $\beta$ -sheet A. There is a large conformational change when substrates are bound and establishment of a nonpolar environment would insure high quantum yield. Conti et al. [46] suggest that the C-terminal region moves to form a cap. The amino-acid residues that we suggest might be involved in the interaction with CoA are found on both sides of the cleft. There are two peptide sequences that are labeled with radioactive thiourea dioxide, a lysine-reagent [14]. Those two peptides are positioned to be part of the active site — one in the C-terminal domain and the other in the N-terminal domain on faces of the cleft that could meet. The Y-339 and S-419 (*Photuris* numbering) suggested above are predicted to be a part of the *P. pyralis* active site containing (*P. pyralis* sequence numbers) S-198, K-206, Y-340, E-344, E-389, Y-401, S-420, G-421, and D-422 [46].

Baldwin [47], in a review of the firefly luciferase structure that reveals a new protein fold, concludes that 'the mystery remains' as to the molecular mechanism of catalysis. The crystals were obtained without bound substrates or other ligands and the suggested active site is based on those amino-acid residues that are conserved among a superfamily of adenylate-forming enzymes.

Elucidation of the amino-acid residues involved in catalyzing the firefly luciferase reactions awaits the results of specific biochemical modification experiments, additional site-directed mutagenesis studies, and further X-ray crystallographic analysis of the protein structure with bound reactants. The large



conformational changes that occur during catalysis suggest that the dynamics of these changes are important in understanding the mechanism. Firefly luciferase differs from the adenylate-forming enzymes used in the structural comparisons described above because luciferase is a monooxygenase and there must be conserved amino-acid residues for that activity. The tools are rapidly becoming available for a better understanding of this bioluminescent process.

### Acknowledgements

This research was supported in part by the Oklahoma Agricultural Experiment Station (Project 1806) and is published with the approval of the director. Thanks to Ms. Ann Williams, Janet Rogers, and Sue Ann Hudiburg of the Sarkeys' Biotechnology Laboratory (Recombinant DNA/Protein Resource Facility) for primer synthesis and automated DNA sequencing. Dr. Don Arnold of the K.C. Emerson Entomology Museum, Oklahoma State University identified firefly species and checked classification schemes. Dr. Ulrich Melcher evaluated the sequencing data and interpretation of the biochemical tree relationships. Dr. Melanie Palmer introduced us to specific molecular biology procedures used with insects and provided materials for practice. Drs. Jerry Devine and Thomas Baldwin provided technical advice on screening clones for luciferase production. The sequences for *Pyrocoelia miyako* and *Hotaria parvula* were provided by Drs. Y. Ohmiya, N. Ohba, H. Toh, and F.I. Tsuji before publication. Dr. Keith Wood provided a correction on our sequence (detected a typographical error that had been entered) and made his unpublished sequences for two *Photuris* luciferases available. Dr. Steve Hartson advised on molecular biology techniques. This paper was read and improved by suggestions from Drs. James Blair, John Cushman, Otis Dermer, Sharon Ford, Robert Matts, Ulrich Melcher, and Keith Wood.

### References

- [1] Seliger, H.H. and McElroy, W.D. (1960) Arch. Biochem. Biophys. 88, 136–141.
- [2] Leach, F.R. (1981) J. Appl. Biochem. 3, 473–517.
- [3] Lundin, A. (1982) in Luminescent Assays (Serio, M. and Pazzagli, M., eds.), pp. 29–45. Raven Press, NY.
- [4] Kricka, L.J. (1988) Anal. Biochem. 174, 14–21.
- [5] Campbell, A.K. (1988) Chemiluminescence: Principles and Applications in Biology and Medicine. Ellis Horwood Ltd., Chichester, 608 pp.
- [6] De Wet, J.R., Wood, K.V., Helinski, D.R. and DeLuca, M. (1985) Proc. Natl. Acad. Sci. USA 82, 7870–7873.
- [7] de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) Mol. Cell. Biol. 7, 725–737.
- [8] Wood, K.V., Lam, Y.A., Seliger, H.H. and McElroy, W.D. (1989) Science 244, 700–702.
- [9] Masuda, T., Tatsumi, M. and Nakano, E. (1989) Gene 77, 265–270.
- [10] Tatsumi, H., Kajiyama, N. and Nakano, E. (1992) Biochim. Biophys. Acta 1131, 161–165.
- [11] Devine, J.H., Kutuzova, G.D., Green, V.A., Ugarova, N.N. and Baldwin, T.O. (1993) Biochim. Biophys. Acta 1173, 121–132.
- [12] Ohmiya, Y., Ohba, N., Toh, H. and Tsuji, F.I. (1995) Photochem. Photobiol. 62, 309–313.
- [13] Sala-Newby, G.B., Thomson, C.M. and Campbell, A.K. (1996) Biochem. J. 313, 761–767.
- [14] Ye, L. (1994) Firefly Luciferase: Modification and Cloning. Dissertation, Oklahoma State University, Stillwater, OK. Order No. AADAA-19525453, Dissertation Abstracts.
- [15] Lloyd, J.B. (1965) Science, 149, 653–654.
- [16] Lloyd, J.B. (1984) Fla. Entomol. 67, 368–376.
- [17] Strause, L.G., DeLuca, M. and Case, J.F. (1979) J. Insect Physiol. 25, 339–348.
- [18] Strause, L.G. and DeLuca, M. (1981) Insect Biochem. 11, 417–422.
- [19] Worley, K.C., Wiese, B.A. and Smith, R.F. (1995) Genome Res. 5, 173–184.
- [20] Henikoff, S. and Henikoff, J.G. (1994) Genomics 19, 97–107.
- [21] Henikoff, S. and Henikoff, J.G. (1991) Nucleic Acids Res. 16, 6565–6572.
- [22] Douglas, S.E. (1994) in Computer Analysis of Sequence Data, Part II (Griffin, A.M. and Griffin, H.G., eds.), Methods in Molecular Biology 25, pp. 181–214. Humana Press, Totowa, NJ.
- [23] Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1994) Nucleic Acids Res. 22, 3441–3444.
- [24] Olson, S.A. (1994) in Computer Analysis of Sequence Data, Part II (Griffin, A.M. and Griffin, H.G., eds.), Methods in Molecular Biology 25, Ch. 17–21. Humana Press, Totowa, NJ.
- [25] Bairoch, A. and Boeckmann, B. (1991) Nucleic Acids Res. 19, 2247–2249.
- [26] Sonnhammer, E.L.L. and Kahn, D. (1994) Protein Sci. 3, 482–492.
- [27] Bairoch, A. and Bucher, P. (1994) Nucleic Acids Res. 22, 3583–3589.
- [28] Bairoch, A. and Boeckmann, B. (1994) Nucleic Acids Res. 22, 3580–3581.

- [29] Herring, P.J. (1987) *J. Biolumin. Chemilumin.* 1, 147–163.
- [30] Jackowski, S., Jackson, P.D. and Rock, C.O. (1994) *J. Biol. Chem.* 269, 2921–2928.
- [31] Toh, H. (1990) *Protein Seq. Data Anal.* 3, 517–521.
- [32] Fulda, M., Heinz, E. and Wolter, F.P. (1994) *Mol. Gen. Genet.* 242, 241–249.
- [33] Fujino, T. and Yamamoto, T. (1992) *J. Biochem.* 111, 197–203.
- [34] Black, P.N., DiRusso, C.C., Metzger, A.K. and Heimert, T.L. (1992) *J. Biol. Chem.* 267, 25513–25520.
- [35] Suzuki, H., Kawarabayashi, Y., Kondo, J., Abe, T., Nishikawa, K., Kimuro, S., Hashimoto, T. and Yamamoto, T. (1990) *J. Biol. Chem.* 265, 8681–8685.
- [36] Pavela-Vrancic, M., Pfeifer, E., van Liempt, H., Schafer, H.-J., Von Doehren, H. and Kleinkauf, H. (1994) *Biochemistry* 33, 6276–6283.
- [37] Maccabe, A.P., van Liempt H., Palissa, H., Unkles, S.E., Riach, M.B.R., Pfeifer, E., Von Doehren, H. and Kinghorn, J.R. (1991) *J. Biol. Chem.* 266, 12646–12654.
- [38] Babbitt, P.C., Kenyon, G.L., Martin, B.M., Charest, H. and Slyvestre, M. (1992) *Biochemistry* 31, 5594–5604.
- [39] Airth, R.L., Rhodes, W.C. and McElroy, W.D. (1958) *Biochim. Biophys. Acta* 27, 519–532.
- [40] Wood, K.V. (1991) in *Bioluminescence and Chemiluminescence: Current Status* (Stanley, P.E. and Kricka, L.J., eds.), pp. 543–546.
- [41] Ford, S.R., Buck, L.M. and Leach, F.R. (1995) *Biochim. Biophys. Acta* 1252, 180–184.
- [42] Wood, K.V. (1995) *Photochem. Photobiol.* 62, 662–673.
- [43] Ford, S.R., Chenault, K.H., Bunton, L.S., Hampton, G.J., McCarthy, J., Hall, M.S., Pangburn, S.J., Buck, L.M. and Leach, F.R. (1996) *J. Biolumin. Chemilumin.* 11, 149–167.
- [44] Leslie, A.G.W. (1990) *J. Mol. Biol.* 213, 168–186.
- [45] Day, P.J., Shaw, W.V., Gibbs, M.R. and Leslie, A.G.W. (1992) *Biochemistry* 31, 4198–4205.
- [46] Conti, E., Franks, N.P. and Brick, P. (1996) *Structure*, 4, 287–298.
- [47] Baldwin, T.O. (1996) *Structure*, 4, 223–228.

## Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening

Ji-HU ZHANG\*, GLENN DAWES†, AND WILLEM P. C. STEMMER\*‡

\*Maxygen, Inc., and †Affymax Research Institute, 3410 Central Expressway, Santa Clara, CA 95051

Communicated by Charles Yanofsky, Stanford University, Stanford, CA, February 26, 1997 (received for review December 26, 1996)

**ABSTRACT** An efficient  $\beta$ -fucosidase was evolved by DNA shuffling from the *Escherichia coli* *lacZ*  $\beta$ -galactosidase. Seven rounds of DNA shuffling and colony screening on chromogenic fucose substrates were performed, using 10,000 colonies per round. Compared with native  $\beta$ -galactosidase, the evolved enzyme purified from cells from the final round showed a 1,000-fold increased substrate specificity for *o*-nitrophenyl fucopyranoside versus *o*-nitrophenyl galactopyranoside and a 300-fold increased substrate specificity for *p*-nitrophenyl fucopyranoside versus *p*-nitrophenyl galactopyranoside. The evolved cell line showed a 66-fold increase in *p*-nitrophenyl fucosidase specific activity. The evolved fucosidase has a 10- to 20-fold increased  $k_{cat}/K_m$  for the fucose substrates compared with the native enzyme. The DNA sequence of the evolved fucosidase gene showed 13 base changes, resulting in six amino acid changes from the native enzyme. This effort shows that the library size that is required to obtain significant enhancements in specificity and activity by reiterative DNA shuffling and screening, even for an enzyme of 109 kDa, is within range of existing high-throughput technology. Reiterative generation of libraries and stepwise accumulation of improvements based on addition of beneficial mutations appears to be a promising alternative to rational design.

Proteins and enzymes with novel functions and properties can be obtained either by searching the largely unknown natural species or by improving upon currently known natural proteins or enzymes. The latter approach may be more suitable for creating properties for which natural evolutionary processes are unlikely to have been selected.

One promising strategy to create such novel properties is by directed molecular evolution. Starting with known natural protein(s), multiple rounds of mutagenesis, functional screening, and amplification can be carried out. When the mutation rate, library size, and selection pressures are properly balanced, the desired phenotype of a protein generally increases with each round (1–8). The advantage of such a process is that it can be used to rapidly evolve any protein, without any knowledge of its structure.

A number of different mutagenesis strategies exist, such as oligonucleotide cassette mutagenesis, point mutagenesis by error-prone PCR or the use of mutator strains, as well as DNA shuffling (1–5, 8). A theoretical approach to choosing a preferred mutagenesis strategy would be to determine the target protein's fitness landscape (9), which is a plot of fitness (on the *y* axis) versus sequence space (on the *x* axis). However, because the sequence space of an average protein of 500 amino acids is  $20^{500}$ , determination of even a fraction of the fitness landscape is a nearly impossible and impractical undertaking.

Because there are just a few fundamental ways to search sequence space, it may be informative to compare the performance of these methods for specific model systems.

Natural genes are thought to have evolved by mutation and recombination within a population of diverse, but highly related, sequences. We suggest that a search algorithm similar to that which slowly created the fitness landscape of a natural protein in the first place is likely to also be the preferred method for further searching this natural sequence landscape (5, 10). This approach is supported by our demonstration of the advantage of recombining mutations (over introduction of point mutations alone) for increasing the activity of a natural  $\beta$ -lactamase protein (2). However, recombination may not always be the best search algorithm. For searching the fitness landscapes of nonnatural sequences under unusual conditions, it is conceivable that a different approach may be more optimal.

We obtain *in vitro* recombination of infrequent point mutations by a PCR-based technique called DNA shuffling (1–5). A pool of closely related sequences is fragmented randomly, and these fragments are reassembled into full-length genes via self-priming PCR and extension in a process we call reassembly PCR (4). This process yields crossovers between related sequences due to template switching. Shuffling allows rapid combination of positive-acting mutations and simultaneously flushes out negative-acting mutations from the sequence pool (Fig. 1). When coupled with effective selection and applied reiteratively, such that the output of one cycle is the input for the next cycle, reiterative DNA shuffling has been demonstrated to be an efficient process for directed molecular evolution (1–3).

In our previous shuffling studies we used selection and/or large libraries (1, 2). Our primary goal in this work was to determine whether detection by screening of libraries of 10,000 clones, a number that is within range of any high throughput screening procedure, would be sufficient to obtain significant enhancement of a minor activity of  $\beta$ -galactosidase, a highly specific and complex enzyme, and at 109 kDa, one of the largest single-chain proteins in *Escherichia coli*. If screening would detect significant improvement, we then would establish that improvements are obtainable by evolution with such small libraries.

*E. coli*  $\beta$ -galactosidase, encoded by *lacZ* (11), is widely used, and its biological function, catalytic mechanism, and molecular structures are well characterized (11–15). It is a tetramer of identical subunits of 1,023 amino acids (13, 16, 17). The crystal structure of  $\beta$ -galactosidase is solved and shows that each subunit forms five structural domains (14). Each active site resides mainly in one subunit, but part of another subunit also is involved (14). The native enzyme hydrolyzes  $\beta$ -galactosyl linkages, such as the  $\beta$ (1, 4)-linkage in its natural disaccharide

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA  
0027-8424/97/944504-6\$2.00/0

PNAS is available online at <http://www.pnas.org>.

Abbreviations: ONPG, *o*-nitrophenyl  $\beta$ -D-galactopyranoside; ONPF, *o*-nitrophenyl  $\beta$ -D-fucopyranoside; PNPG, *p*-nitrophenyl  $\beta$ -D-galactopyranoside; PNPF, *p*-nitrophenyl  $\beta$ -D-fucopyranoside; X-Fuc, 5-bromo-4-chloro-3-indolyl  $\beta$ -D-fucopyranoside.

‡To whom reprint requests should be addressed. e-mail: [maxygen@maxygen.com](mailto:maxygen@maxygen.com).

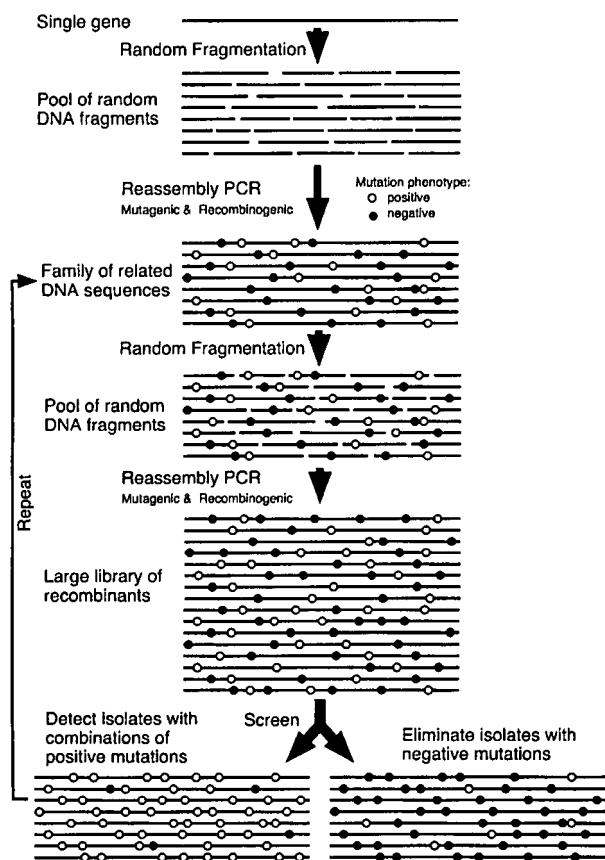


FIG. 1. Schematic illustration of the DNA shuffling process used in the present study.

substrate, lactose. The native  $\beta$ -galactosidase is known to be highly specific for  $\beta$ -D-galactosyl substrates. A multistep model of the reaction was proposed (18, 19) based on kinetic studies of the native enzyme for *o*-nitrophenyl  $\beta$ -D-galactopyranoside (ONPG), *p*-nitrophenyl  $\beta$ -D-galactopyranoside (PNPG), and other substrates and substrate analogs. The native  $\beta$ -galactosidase acts only weakly on  $\beta$ -D-fucosyl moieties (18–20) and does not act on most substrate analogs.

## MATERIALS AND METHODS

*E. coli*  $\beta$ -galactosidase (EC 3.2.1.23) and the galactosyl and fucosyl substrates 5-bromo-4-chloro-3-indolyl  $\beta$ -galactopyranoside (X-Gal), PNPG, ONPG, 5-bromo-4-chloro-3-indolyl  $\beta$ -D-fucopyranoside (X-Fuc), *p*-nitrophenyl  $\beta$ -D-fucopyranoside (PNPF), and *o*-nitrophenyl  $\beta$ -D-fucopyranoside (ONPF) were purchased from Sigma. Plasmid pCH110 containing a *lacZ* gene was from Pharmacia. *E. coli* strain TB1 was a gift from Charles Roessner of Texas A&M University.

**Construction of Plasmid p18lacZ.** A 3.8-kb *Hind*III and *Bam*HI restriction nuclease fragment from pCH110 containing a *lacZ* gene (codon 8 fused to a short N-terminal peptide) and the *gpt* promoter region (21) was subcloned into the *Hind*III and *Bam*HI sites of vector p18-sfi-kan-sfi vector, a 2.3-kb pUC18 derivative in which the ampicillin gene is replaced by a kanamycin phosphotransferase gene (2). The resulting plasmid, named p18lacZ, was used for DNA shuffling. DNA fragments of 50–200 bp were used and reassembled as described previously (1, 2). The PCR primers for amplification of the reassembled genes were AGCGC-CCAATACGCAAACCGCCTCTCCCCGCGGTTGGCC (forward) and CTATGCGGCATCAGAGCAGATTGTACT-

GAGAGTGCACCAT (reverse), located on either side of the *Bam*HI and *Hind*III fragment. The reassembled gene was digested with restriction enzymes *Hind*III and *Bam*HI and ligated back into the P18-sfi-kan-sfi vector. The ligation mixture was electroporated into *E. coli* TB1 competent cells and plated out on Luria-Bertani plates (150 mm) with 40  $\mu$ g/ml kanamycin and 2 mg/plate of the X-Fuc substrate (22). The plates were incubated 12 to 24 hr at 37°C. The resulting kanamycin-resistant transformants were visually screened for the intensity of the blue color. The 20–40 colonies with the most intense blue color were picked from about 10,000 transformants of each round and used for the next round of DNA shuffling. Seven rounds of DNA shuffling and screening were carried out. The best clone from the final screening round, called evolved  $\beta$ -fucosidase, was characterized in detail.

**Enzyme Purification.** For purification of the native  $\beta$ -galactosidase and the evolved  $\beta$ -fucosidase, a histidine tag (His<sub>6</sub>) was fused to the N terminus of both enzymes by PCR with two primers [5'-(P)CATCACCATCACCACCATATCGTCAC-CTGGGACATGT and 5'-(P)GTATTTTTCGCTCATGT-GAA] in a standard PCR. The histidine-tagged native and evolved enzymes were purified from overnight TB1 cell cultures harboring the corresponding plasmid (23). The crude cell extract, in 50 mM phosphate (pH 7.0) with 100 mM NaCl and 0.2 mM of phenylmethylsulfonyl fluoride protease inhibitor was passed through a 20-ml Ni-nitrilotriacetic acid agarose (Qiagen) column. The bound protein was stepwise-eluted with the same buffer containing 5 mM, 10 mM, 25 mM, and 100 mM imidazole. The active fractions from the metal affinity column were desalted and loaded on a DEAE column in 20 mM Tris (pH 7.5), followed by elution with a 0 to 1 M NaCl gradient. The active fractions were concentrated and loaded on a Superose 12 gel filtration column in an FPLC protein purification unit (Pharmacia). SDS/PAGE analysis (data not shown) showed that the native galactosidase and the evolved fucosidase were greater than 90% pure.

**Enzyme Kinetics.**  $\beta$ -Galactosidase activity was assayed using the synthetic chromogenic substrates ONPG and PNPG.  $\beta$ -Fucosidase activity was assayed using chromogenic fucosyl substrates ONPF and PNPF. Enzyme assays were performed at 25°C and pH 7.0 in 30 mM *N*-tris(hydroxymethyl)methylaminoethanesulfonic acid with 1 mM MgCl<sub>2</sub> and 150 mM NaCl. The absorbance change at 420 nm was recorded with time, and product formation was quantitated using the absorption extinction coefficient (2.65 mM<sup>-1</sup>·cm<sup>-1</sup> for *o*-nitrophenol and 6.7 mM<sup>-1</sup>·cm<sup>-1</sup> for *p*-nitrophenol). For kinetic parameter measurements, the initial velocity  $V_0$  (when less than 10% of the substrate was converted into product) was determined with varied substrate concentrations. The values of  $V_{max}$  and  $K_m$  were calculated using the simple weighting method of Cornish-Bowden (24). The  $V_{max}$  values were converted to  $k_{cat}$  values, the turnover number per active site, by normalizing for the enzyme concentrations by the molecular mass of the monomer. The  $K_m$  and  $k_{cat}$  values of the wild-type  $\beta$ -galactosidase for ONPF could not be determined directly because of the low activity on this substrate. The  $k_{cat}/K_m$  value had to be estimated from the enzyme dilution factor required for the native enzyme to generate the same amount of *o*-nitrophenol product from ONPG after the same period of time (usually several hours) and from the  $k_{cat}/K_m$  value of the wild-type enzyme on ONPG.

**Sequencing of the Evolved *lacZ* Gene.** The 3.8-kb DNA fragment encoding the evolved  $\beta$ -galactosidase and its flanking regions was sequenced in both forward and reverse directions with 20 primers using an Applied Biosystems 391 DNA sequencer.

## RESULTS AND DISCUSSION

**Strategy for Evolving  $\beta$ -Galactosidase.** The primary goal of the experiment was to determine if a substantial enhancement

in the specificity and/or activity of a large model enzyme could be obtained by reiterative screening of libraries of a size (10,000 clones) that is routinely accessible by high throughput detection assays. No structural information was used in the design of the experiment, but the structure of  $\beta$ -galactosidase is useful for interpretation of results.

**Screening for Improved Fucosidase Activity.** The 3.8-kb DNA fragment of p18LacZ containing the *lacZ* gene was shuffled as described previously (1–3), and the reassembled genes were digested with restriction enzymes (*Hind*III and *Bam*HI) and ligated back into the vector p18-sfi-kan-sfi. The initial diversity was introduced into the native *lacZ* gene by random point mutagenesis, which occurs by shuffling of small fragments (1, 2). We previously showed that shuffling with 10- to 50-bp fragments resulted in a 0.7% rate of point mutation. Here we used fragments of 50 to 200 bp, which results in a much lower rate of point mutation, resulting in inactivation of approximately 20% of the clones. X-Fuc was chosen as the indicator substrate for the plate assay because of the nondiffusible nature of the colored product and the high sensitivity (22). After each round of DNA shuffling, 10,000 kanamycin-resistant transformants, growing on plates supplemented with X-Fuc, were visually screened for enhanced blue color formation. About 2–5% of the transformant colonies in each round showed colonies that were more highly blue-colored than the bulk of the population. The 20–40 bluest colonies (0.2–0.4%) were picked at each round, individually verified to be more active than the pool from the previous round by plate assays, and then used as a pool for the source of DNA to initiate the next round of DNA shuffling. This number of colonies was chosen as a compromise between obtaining too little diversity (<10 colonies) and obtaining suboptimal selection pressure ( $\gg$ 100 colonies), which could limit the rate of improvement. In the seventh round of DNA shuffling some colonies developed a deep blue color after overnight growth (Fig. 2). One mutant from this seventh and final round of shuffling showed a 66-fold increase in fucosidase activity on 1 mM PNPF (Fig. 3).

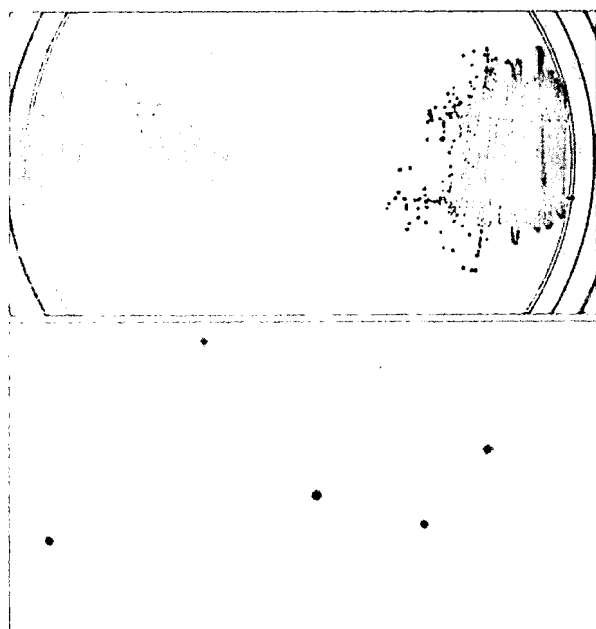


FIG. 2. *E. coli* TB1 cells expressing the native  $\beta$ -galactosidase (white colonies, Upper Left) and the evolved fucosidase of the seventh round (blue colonies, Upper Right) after overnight growth on an Luria-Bertani plus kanamycin plate supplemented with 0.1 mM X-Fuc. (Lower) The results of plating a deliberate mixture of the two types of colonies.

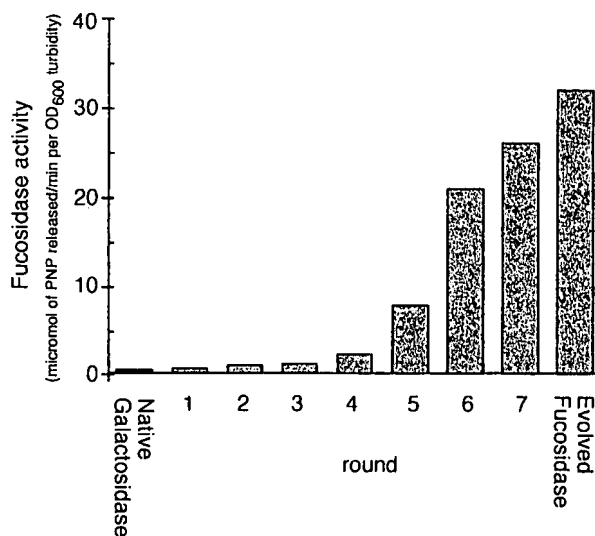


FIG. 3. Whole cell fucosidase activity on PNPF of the pool of colonies selected after each round of DNA shuffling. Rounds 1–7 are pools of colonies. Also shown are the activity of cells expressing the native  $\beta$ -galactosidase and cells expressing the evolved  $\beta$ -fucosidase, both measured as whole cell activity of single clones. The evolved fucosidase is the single-best colony selected after quantitative comparison of the 24 best colonies from the pool of colonies obtained after shuffling round 7. For assay conditions, see *Materials and Methods*.

**Kinetics.** After the final round of selection, (His)<sub>6</sub> tags were added to the foreign N terminus of the native  $\beta$ -galactosidase and the evolved  $\beta$ -fucosidase enzymes. Both enzymes were purified, and the kinetic constants of each enzyme on the synthetic chromogenic substrates ONPG, ONPF, PNPG, and PNPF were determined (Table 1). For PNPF, the  $K_m$  value of the evolved fucosidase is decreased by 20-fold from the  $K_m$  of wild-type  $\beta$ -galactosidase on the same substrate. The  $k_{cat}$  value is decreased about 2-fold. The  $k_{cat}/K_m$  values thus are increased about 10-fold in the evolved  $\beta$ -fucosidase. The activity of the wild-type enzyme on ONPF was very low and accurate  $K_m$  and  $k_{cat}$  values could not be obtained. By comparing the relative reaction rates of the

Table 1. Kinetic constants for the native and evolved enzymes

Substrate	Kinetic constant	Native galactosidase	Evolved fucosidase
PNPG	$k_{cat}$ , s <sup>-1</sup>	268	30.9
	$K_m$ , mM	0.04	0.18
	$k_{cat}/K_m$ , mM <sup>-1</sup> s <sup>-1</sup>	6,700	172
PNPF	$k_{cat}$ , s <sup>-1</sup>	209	96.6
	$K_m$ , mM	31	1.5
	$k_{cat}/K_m$ , mM <sup>-1</sup> s <sup>-1</sup>	6.7	64.4
Specificity	$(k_{cat}/K_m)_{PNPG}$	1,000	2.7
	$(k_{cat}/K_m)_{PNPF}$		
ONPG	$k_{cat}$ , s <sup>-1</sup>	765	14.5
	$K_m$ , mM	0.11	0.11
	$k_{cat}/K_m$ , mM <sup>-1</sup> s <sup>-1</sup>	6,950	132
ONPF	$k_{cat}$ , s <sup>-1</sup>	—	24.1
	$K_m$ , mM	—	0.55
	$k_{cat}/K_m$ , mM <sup>-1</sup> s <sup>-1</sup>	(2)*	43.9
Specificity	$(k_{cat}/K_m)_{ONPG}$	3,200	3.0
	$(k_{cat}/K_m)_{ONPF}$		

The native galactosidase and the evolved fucosidase were purified, and the enzymes were assayed on four different substrates.

\*The  $k_{cat}/K_m$  value for the native galactosidase on ONPF was estimated to be about 2 mM<sup>-1</sup>s<sup>-1</sup> by measuring the hydrolysis rate relative to that of ONPG.

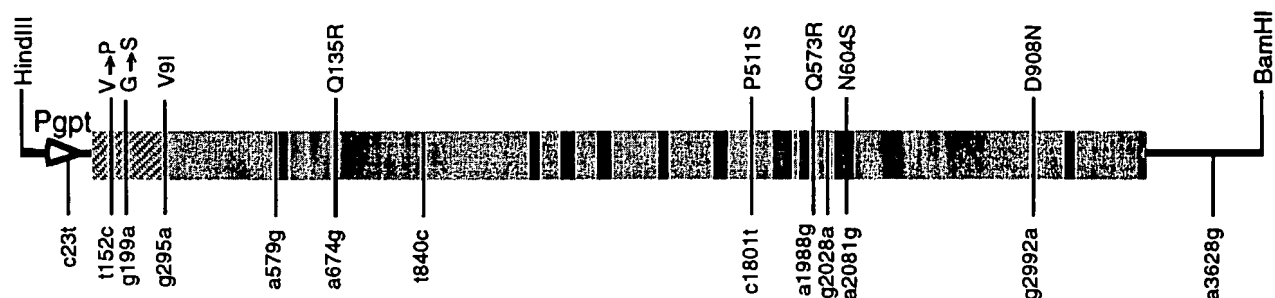


FIG. 4. Nucleotide substitutions in the evolved fucosidase gene. The predicted amino acid changes are shown above the gene by the single-letter denotation, numbered according to the wild-type  $\beta$ -galactosidase sequence (17). Amino acid changes in the N-terminally fused peptide region (hatched area) are indicated by small vertical arrows. Mutations that do not result in amino acid changes are shown below the gene, numbered starting at the *Hind*III site, as in the parental plasmid pCH110. The gpt promoter is indicated by a thick arrow. The positions of the known active site residues of the wild-type  $\beta$ -galactosidase are indicated by black bars.

wild-type enzyme on ONPF and ONPG (at the same enzyme and substrate concentrations), the  $k_{cat}/K_m$  for ONPF was estimated, assuming that the  $k_{cat}/K_m$  value is a second order rate constant. The  $k_{cat}/K_m$  values on ONPF were increased at least 20-fold in the evolved  $\beta$ -fucosidase. These increases in fucosidase activity were accompanied by decreases in galactosidase activity. For the substrates PNPG and ONPG, the  $k_{cat}/K_m$  is decreased 40-fold and 50-fold, respectively. These kinetic parameter changes suggest that the substrate binding pocket in the evolved  $\beta$ -fucosidase is different from that of the wild-type  $\beta$ -galactosidase.

The native enzyme is highly specific for hydrolyzing galactosyl rather than fucosyl substrates. The  $k_{cat}/K_m$  values we determined for PNPG and PNPF differ by about 1,000-fold, and for ONPG and ONPF the values differ by more than 3,000-fold (Table 1). The values we determined for the native  $\beta$ -galactosidase on ONPG, PNPG, and PNPF are in between the values reported previously (18, 20). The substrate specificity changed dramatically from the native  $\beta$ -galactosidase to the evolved  $\beta$ -fucosidase. For the evolved  $\beta$ -fucosidase the  $k_{cat}/K_m$  values for substrates PNPG and PNPF differ 2.7-fold and for substrates ONPG and ONPF the  $k_{cat}/K_m$  values differ

3-fold. Therefore, the relative substrate specificity for fucosyl substrates, from the native to the evolved enzyme, is increased 1,000-fold for the *o*-nitrophenyl substrates and 300-fold for the *p*-nitrophenyl substrates. The substrate specificity change was further supported by inhibition of the enzymatic activity by isopropyl  $\beta$ -D-thiogalactopyranoside, a  $\beta$ -galactosidase substrate analog and a competitive inhibitor of galactosyl substrates. The  $K_i$  values increased by one order of magnitude from the wild-type enzyme to the evolved  $\beta$ -fucosidase, from 0.1 mM to 0.9 mM. The changes in  $K_m$  values for the galactosyl substrates showed the same trend, because they either increased severalfold or stayed the same. These results imply that the substrate binding affinity of the evolved  $\beta$ -fucosidase is substantially increased for fucosyl substrates and decreased for galactosyl substrates, and hence the substrate binding pocket is likely to be significantly modified.

**DNA Sequence.** The DNA sequence of the evolved fucosidase gene showed 13 nucleotide substitutions of which 11 were in the coding region. Six of the mutations are predicted to cause amino acid changes in the translated  $\beta$ -galactosidase sequence. Two additional mutations are predicted to cause amino acid changes in the N-terminal fusion peptide (Fig. 4).

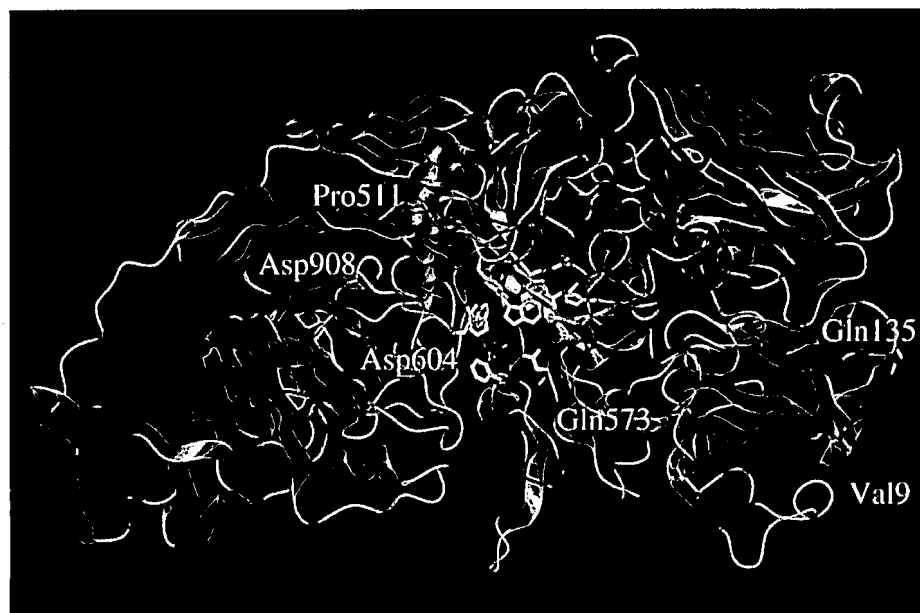


FIG. 5. Ribbon representation of the *E. coli*  $\beta$ -galactosidase subunit structure (14). The CNO atoms of the six amino acid mutations that conferred the fucosidase activity are shown with stick representation. Two mutations in the active site (Asp604 and Gln573) are shown in red. Two mutations in close proximity of the active site (Pro511 and Asp908) are shown in magenta. Two mutations far away from the active site and on the protein surface (Val9 and Gln135) are shown in green. The rest of the substrate binding and active site residues are shown in yellow.

All 13 nucleotide changes were base transitions between purines and/or between pyrimidines, which usually are more frequent than transversions.

One major advantage of *in vitro* evolution of enzymes over the structural modeling approach is that only minimal information is required for improving the desired phenotype. At each round of our experiment, only colonies with increased fucosidase activity were pooled and used for the next round of DNA shuffling and screening. Although both positive-acting mutations and neutral mutations may accumulate in the evolved fucosidase *lacZ* gene in each round, we expect that neutral mutations generally do not survive multiple rounds of shuffling and screening due to a backcrossing effect exerted by the consensus sequence (1, 2), combined with the lack of a selective advantage of the neutral mutations. Therefore, only mutations that somewhat contribute to the improved fucosidase activity are likely to accumulate in the evolved fucosidase. While we have not determined the effect of the separate mutations by site-specific mutational studies, we can predict what roles some of the mutations may play based on the three-dimensional structure of the parental  $\beta$ -galactosidase (ref. 14; Fig. 5) and the sophisticated kinetic models based on previous mutations and kinetic analysis of purified proteins (18, 20, 25, 26). Among the six amino acid changes in the  $\beta$ -galactosidase sequence, none appear directly involved in the inter-subunit contact. Three mutations (Pro511Ser, Gln573Arg, and Asn604Ser) are located in domain 3 (residues 334–627) of the wild-type *E. coli*  $\beta$ -galactosidase (14). Domain 3 in the native protein contains most of the amino acids that form the substrate binding pocket (ref. 14; Fig. 5). Asn604 is one of the amino acids forming this substrate binding pocket in the protein (14), and this residue is conserved in several other known  $\beta$ -galactosidase sequences (25, 27–29), except the evolved galactosidase gene (*ebgA*) of *E. coli* (30). In our evolved fucosidase, Asn604 is replaced by Ser. This mutation presumably affects the enzyme's substrate specificity. All the other mutations found in the evolved  $\beta$ -fucosidase enzyme do not directly affect the active site and substrate binding pocket residues, and therefore they may have no effect or may only subtly change the conformation of the active site and substrate specificity. Gln573, substituted by Arg in the evolved fucosidase, is in close proximity to the substrate binding pocket (Fig. 5). The mutation Pro511Ser is also close to the active site and substrate binding pocket (Fig. 5). These two mutations are likely to affect the enzyme's active site. Asp908Asn is also close to the active site and may also affect the activity. Additional important catalytic residues of the active site, such as Glu461, Met502, Tyr503, and Glu537 (23, 26, 31), however, are unchanged in the evolved  $\beta$ -fucosidase, implying that the catalytic mechanism of the evolved enzyme remained the same. Therefore the evolved  $\beta$ -fucosidase seems to have only adjusted to fit the fucosyl substrate or its transition state better than the wild-type  $\beta$ -galactosidase does. In addition, one of the nucleotide mutations outside the structural gene (*c231*) is very close to the *gpt* promoter region and could affect transcription (Fig. 4). This mutation, along with the two amino acid mutations in the N-terminal fusion peptide (Fig. 4), may influence the expression level of the protein. Indeed, we found that the evolved  $\beta$ -fucosidase enzyme was expressed at least 2- to 3-fold higher than the wild-type enzyme (data not shown). The mutations Val91Ile and Gln135Arg are far away from the active site and near the surface of the protein (Fig. 5), and may not have any significant effect on the enzymatic activity. The analysis of mutations obtained by molecular evolution of proteins provides a new tool for studying structure–function relationships. However, the real utility of DNA shuffling is the

ability to rapidly improve enzyme functions without the need to delineate the myriads of complex molecular mechanisms.

There are several possible applications for the evolved  $\beta$ -fucosidase. One is as a novel reporter for  $\beta$ -D-fucosyl substrates, in addition to the widely used *lacZ* gene reporter. The advantage of using the novel enzyme is the low endogenous background of  $\beta$ -fucosidase activity because, unlike  $\alpha$ -fucosidases,  $\beta$ -fucosidases are uncommon in nature. This well expressed fucosidase also could be used for the production of fucosyl adducts or for disaccharide synthesis by transglycosylation or reversal of the hydrolysis reactions, because analogous applications already have been demonstrated for the wild-type  $\beta$ -galactosidase (32, 33). Some of these applications may require further evolution of the fucosidase for the specific reaction. The present data suggests that it is reasonable to attempt to obtain such improvements by DNA shuffling and screening of libraries of modest size.

We thank A. Cramer for technical assistance and Drs. F. H. Arnold, R. E. Huber, and P. Schatz for useful discussions.

1. Stemmer, W. P. C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751.
2. Stemmer, W. P. C. (1994) *Nature (London)* **370**, 389–391.
3. Cramer, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996) *Nat. Biotechnol.* **14**, 315–319.
4. Stemmer, W. P. C. (1996) in *The Encyclopedia of Molecular Biology* (VCH, New York), pp. 447–457.
5. Stemmer, W. P. C. (1995) *Bio/Technology* **13**, 549–553.
6. You, L. & Arnold, F. H. (1995) *Protein Eng.* **9**, 77–83.
7. Chen, K. & Arnold, F. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5618–5622.
8. Moore, J. C. & Arnold, F. H. (1996) *Nat. Biotechnol.* **14**, 458–467.
9. Kauffman, S. A. (1993) *The Origins of Order* (Oxford Univ. Press, New York).
10. Stemmer, W. P. C. (1995) *Science* **270**, 1510.
11. Beckwith, J. (1996) in *Escherichia coli and Salmonella*, ed., Neidhardt, F. (Am. Soc. Microbiol., Washington, DC), pp. 1227–1231.
12. Zabin, I. & Fowler, A. V. (1980) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 89–122.
13. Wallenfels, K. & Weil, R. (1972) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), pp. 618–663.
14. Jacobson, R. E., Zhang, X.-J., DuBose, R. F. & Matthews, B. W. (1994) *Nature (London)* **369**, 761–766.
15. Huber, R. E., Gupta, M. N. & Khare, S. K. (1994) *Int. J. Biochem.* **26**, 309–318.
16. Fowler, A. & Zabin, I. (1978) *J. Biol. Chem.* **253**, 5521–5525.
17. Kalnins, A., Otto, K., Ruther, U. & Muller-Hill, B. (1983) *EMBO J.* **2**, 593–597.
18. Roth, N. J. & Huber, R. E. (1996) *J. Biol. Chem.* **271**, 14296–14301.
19. Huber, R. E. & Gaunt, M. T. (1983) *Arch. Biochem. Biophys.* **220**, 263–271.
20. Wallenfels, K., Lehmann, J. & Malhotra, O. P. (1960) *Biochem. Z.* **333**, 209–215.
21. Hall, C. V., Jacob, P. E., Ringold, G. M. & Lee, F. (1983) *J. Mol. Appl. Genet.* **2**, 101–109.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
23. Cupples, C. G., Miller, J. H. & Huber, R. E. (1990) *J. Biol. Chem.* **265**, 5512–5518.
24. Cornish-Bowden, A. (1976) *Principles of Enzyme Kinetics* (Butterworth, London), pp. 168–189.
25. Buvinger, W. E. & Riley, M. (1985) *J. Bact.* **163**, 850–857.
26. Gebler, J. C., Aebersold, R. & Withers, S. G. (1992) *J. Biol. Chem.* **267**, 11126–11130.
27. Schroeder, C. J., Robert, C., Lenzen, G., McKay, L. L. & Mercenier, A. (1991) *J. Gen. Microbiol.* **137**, 369–380.

28. Schmidt, B. F., Adams, R. M., Requadt, C., Power, S. & Mainzer, S. E. (1989) *J. Bacteriol.* **171**, 625–635.
29. Hancock, K. R., Rockman, E., Young, C. A., Pearce, L., Maddox, I. S. & Scott, D. B. (1991) *J. Bacteriol.* **173**, 3084–3095.
30. Stokes, H. W., Betts, P. W. & Hall, B. G. (1985) *Mol. Biol. Evol.* **2**, 469–477.
31. Ring, M. & Huber, R. E. (1990) *Arch. Biochem. Biophys.* **283**, 342–350.
32. Hedbys, L., Larsson, P. O., Mosbach K. & Svensson, S. (1984) *Biochem. Biophys. Res. Commun.* **123**, 8–15.
33. Huber, R. E. & Hurlburt, K. L. (1986) *Arch. Biochem. Biophys.* **246**, 411–418.





# UNITED STATES PATENT AND TRADEMARK OFFICE

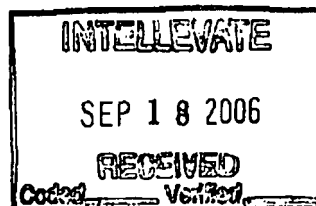
UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

T\_SLW11397

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/645,706	08/24/2000	Keith V. Wood	341.005US1	3329
21186	7590	09/13/2006	EXAMINER	
SCHWEGMAN, LUNDBERG, WOESSNER & KLUTH, P.A. P.O. BOX 2938 MINNEAPOLIS, MN 55402			PROUTY, REBECCA E	
			ART UNIT	PAPER NUMBER
			1652	

DATE MAILED: 09/13/2006

Please find below and/or attached an Office communication concerning this application or proceeding.



Schwegman, Lundberg,  
Woessner & Kluth, P.A.

SEP 18 2006

RECEIVED

<b>Office Action Summary</b>	Application No. 09/645,706	Applicant(s) WOOD ET AL.	
	Examiner Rebecca E. Prouty	Art Unit 1652	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) See Continuation Sheet is/are pending in the application.  
     4a) Of the above claim(s) 64 is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1,3-6,9,11,12,15,18,20,21,24-39,41-45,47,60,67,69-71,74,76-78,80-88 and 90-96 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
     a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |   |  |
|---|--|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. ____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)  | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)            |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date <u>7/06</u> . | 6) <input type="checkbox"/> Other: ____  |

**Continuation of Disposition of Claims:** Claims pending in the application are 1,3-6,9,11,12,15,18,20,21,24-39,41-45,47,60,64,67,69-71,74,76-78,80-88 and 90-96.

Art Unit: 1652

Claims 2, 7, 8, 10, 13, 14, 16, 17, 19, 22, 23, 40, 46, 48-59, 61-63, 65, 66, 68, 72, 73, 75, 79, and 89 have been canceled. Claims 1, 3-6, 9, 11, 12, 15, 18, 20, 21, 24-39, 41-45, 47, 60, 64, 67, 69-71, 74, 76-78, 80-88, 90-94 and newly presented claims 95-96 are still at issue and are present for examination.

Claim 64 remains withdrawn from further consideration pursuant to 37 CFR 1.142(b), as being drawn to a nonelected species, there being no allowable generic or linking claim. Applicant timely traversed the restriction (election) requirement in the response filed 11/18/02.

Applicants' arguments filed on 6/22/06, have been fully considered. Rejections and/or objections not reiterated from previous office actions are hereby withdrawn.

Claims 90, 95, and 96 are objected to because of the following informalities: the word "to" needs to be inserted following "selected" in the phrase "codons are selected reduce the number of identified sequences or sites". Appropriate correction is required.

Claims 1, 3-6, 9, 11, 12, 15, 18, 20, 21, 24-39, 41-45, 47, 60, 67, 69-71, 74, 76-78, 80-83, 85-88, and 90-96 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for

failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claims 1 (from which claims 3-6, 9, 11, 12, 15, 20, 21, 24-39, 41-45, 60, 69, 70 81, 86, and 90 depend), 47 (from which claims 71, 82, and 87 depend), 67 (from which claims 69, 70, 81, 88 and 95 depend), 74 (from which claims 76, 77, 81, 88 and 96 depend), and 78 (from which claims 80, 82, and 87 depend) are vague and indefinite in the recitation of "a reduced number of a combination of mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and/or prokaryotic 5' noncoding regulatory sequences", "wherein the mammalian transcription factor binding sequences are present in a database of transcription factor binding sequences" and "known mammalian transcription factor binding sequences". The rejection was described in the previous Office Action.

Applicants argue that the terms "transcription factor binding sequences", "intron splice sites", "poly(A) addition sites" and "prokaryotic 5' noncoding regulatory sequences" are conventional in the art and argue that these terms are in fact used in the reference cited by the examiner in the 103 rejection. This is acknowledged. However, in the art these terms define a group of sequences related by function. The art does not define clearly **what** sequences are included in the

group. Since applicants invention requires a skilled artisan to **quantify** the number of such sequences it is imperative that the artisan know explicitly what sequences are to be included and what sequences are not so one can in fact count them. While the art clearly defines **some** specific sequences which fall into each group (for example AAUAAA as a polyadenylation sequence) many other sequences may have the same function and not all such sequences are known and taught by the art.

With regard to calculating the number of mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences, point to Example 1 of the specification and the declaration of Dr. Wood submitted with the instant response and argue that both evidence that contrary to the Examiner's assertion, the calculation of the number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences in a particular sequence is possible. However, it has never been the examiner's contention that given a clear set of sequences to be searched that calculation of the number was not possible, but that the claims are indefinite absent a clear definition of what sequences are encompassed by these terms. In Example 1 of the specification it is clearly set forth that transcription factor

Art Unit: 1652

binding sequences were mammalian sequences identified in the SITE table of TRANSFAC database version 3.2 having a minimum length of 5 nucleotides and a minimum log-likelihood (as defined in the spec.) of 10, intron splice sites were AGGTRAGT, AGGTRAG or YNCAGG, poly(A) addition sites were AATAAA, and prokaryotic 5' noncoding regulatory sequences were TATAAT or either of AGGA or GGAG paired with an ATG codon within 12 or fewer bases 3' to said sequence. Each of these is a clearly identifiable and defined set of sequences, presuming that the SITE table of the TRANSFAC database version 3.2 is obtainable. Similarly in the search discussed in the declaration of Dr. Wood, transcription factor binding sites are defined as mammalian sequences identified in the SITE table of TRANSFAC database version 4.0 having a minimum length of 5 nucleotides and a minimum log-likelihood (as defined in the TESS software literature attached) of 10. This is a clearly identifiable and defined set of sequences, presuming that the SITE table of TRANSFAC database version 4.0 is obtainable. However, none of applicants claims is limited in a similar fashion. It is suggested that applicants limit the claims to the sites used in Example 1 of the specification and submit a copy of the SITE table of version 3.2 of the TRANSFAC database.

Claim 18 as amended is confusing in the recitation "or the complement thereof which encodes a luciferase" as the complement does not encode a luciferase.

Claims 47 and 83 as amended are indefinite in the recitation of "corresponding wild type nucleic acid sequence" as it is unclear to what sequences this must refer. Is this limited to the wild type sequences from which SEQ ID NOS: 9, 18, 297, and 301 were derived (i.e., LucPplyG, SEQ ID NO:1) or to any wild type beetle luciferase gene?

Claim 83 is confusing in the recitation of "hybridizes under medium stringency hybridization conditions to SEQ ID NO:22 (Rluc-final) ... and comprises an open reading frame encoding a luciferase with 90% amino acid sequence identity to a beetle luciferase" as SEQ ID NO:22 (Rluc-final) is a variant of *Renilla* luciferase which is not a beetle luciferase and in view of the lack of similarity of *Renilla* luciferase with beetle luciferases, a polynucleotide which hybridizes to SEQ ID NO:22 could not encode a luciferase with 90% amino acid sequence identity to a beetle luciferase. While applicants amended claim 47 to address this problem they did not amend claim 83. As such this rejection is maintained for claim 83.

Claims 1, 3-6, 9, 11, 12, 15, 20-21, 24-33, 35-39, 41-45, 47, 60, 67, 69, 70, 81-82, 86-88, and 90-95 are rejected under



Art Unit: 1652

35 U.S.C. 112, first paragraph, because the specification, while being enabling for (1) a variant of a parent DNA molecule encoding a reporter polypeptide identical to a reporter polypeptide encoded by said parent DNA, having more than 25% of the codons altered and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and promoter sequences than a mammalian codon optimized variant of the parent nucleic acid, (2) a variant of a parent DNA molecule encoding a luciferase having 90% identity to the polypeptide encoded by SEQ ID NO:2 and having more than 25% of the codons altered and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and 5' noncoding regulatory sequences than a mammalian codon optimized variant of SEQ ID NO:2 or (3) to any nucleic acid which will hybridize to SEQ ID NO:9 under high stringency conditions and encode a polypeptide having luciferase activity, does not reasonably provide enablement for any variant DNA molecules encoding any reporter polypeptide having at least 90% identity to a wild type reporter polypeptide, having more than 25% of the codons altered and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and 5' noncoding regulatory sequences than a mammalian codon optimized version of the parent nucleic

Art Unit: 1652

acid or to any nucleic acid which will hybridize to SEQ ID NO:9 under medium stringency conditions. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to use the invention commensurate in scope with these claims. The rejection is explained in the previous Office Action.

Applicants first state that it is unclear how Applicant's specification teaches one of skill in the art how to make and use a variant of a parent DNA molecule encoding a reporter polypeptide identical to a reporter polypeptide encoded by said parent DNA, having more than 25% of the codons altered and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and promoter sequences if the art worker would not recognize or understand sequences that are mammalian transcription factor binding sequences, intron splice sites, poly(A) addition sites and prokaryotic 5' noncoding regulatory sequences. Applicants are in fact correct that, if a skilled artisan cannot clearly identify these sites, he can not practice the invention as taught. However, for the instant rejection the claims were examined as best possible ignoring this problem (as it is clearly addressed by the rejection above) in the interest of compact prosecution, such that all possible problems could be

Art Unit: 1652

identified concurrently. The instant rejection would be maintained even if the claims clearly identified all such sites for the reasons presented.

Next applicants argue that with respect to reporter polypeptides, such as GFP, beetle luciferase, GUS, CAT, and beta-lactamase, applicant has provided evidence that it is well within the skill of the art to introduce substitutions into various reporter proteins and yield a variant protein with the activity of the corresponding wild-type reporter protein. However, it is noted that the evidence applicants refer to is available for specific GFPs, beetle luciferases, GUS or CAT enzymes, and beta-lactamases but each of these groups of reporter polypeptides includes vast numbers of proteins which are not well characterized and often substantially different from those taught in the art. For example there are many different luminescent beetle species but only a few firefly and click beetle luciferases are well characterized in the art and even these enzymes differ from each other enormously. The rejected claims are not limited the nucleic acids encoding reporters exhibiting high similarity to only those reporters which are well characterized (Note claims that are so limited such as claims 18, 71, 74, 76-78, 80, 83-85, and 96 are not rejected).

Finally applicants argue that one of skill in the art in possession of applicant's specification is readily able to determine whether a variant nucleic acid molecule hybridizes under medium stringency conditions to Applicant's synthetic polynucleotides and has an open reading frame encoding a beetle luciferase polypeptide which has at least 90% amino acid sequence identity to a luciferase encoded by a corresponding wild type nucleic acid sequence, and having a reduced number of transcription factor binding sequences, intron splice sites, poly(A) addition sites and 5' noncoding regulatory sequences than a mammalian codon optimized version of the parent nucleic acid. However, while methods of determining if any individual sequence would have the properties described are well known in the art, methods of determining which sequences from the virtually infinite genus of sequences capable of hybridizing to under medium stringency conditions to the recited nucleic acids and encoding a protein having 90% identity to any beetle luciferase actually are within the scope of the instant claims (i.e., encode a luciferase protein) beyond just making and testing all possibilities are not provided. While enablement is not precluded by the necessity for routine screening, if a large amount of screening is required, the specification must provide a reasonable amount of guidance with respect to the direction in

Art Unit: 1652

which the experimentation should proceed. Such guidance has not been provided in the instant specification.

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Claims 1, 3-6, 9, 11, 12, 15, 20, 21, 24-39, 41-45, 60, 67, 69, 70, 81, 86 and 90-95 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sherf et al. (US Patent 5,670,356) in view of Zolotukhin et al. (US Patent 5,874,304), Donnelly et al. (WO 97/47358), Pan et al., Cornelissen et al. (US Patent 5,952,547), and Hey et al. (US Patent 6,169,232). The rejection is explained in the previous Office Action.

Applicants argue that the combination of references does not disclose or suggest Applicant's invention as each reference discloses a different way to modify the coding sequence of a

Art Unit: 1652

different gene to increase expression. This is not persuasive because each of these references is drawn to methods of increasing the expression of a gene in a desired host by altering the sequence of the nucleic acid but not the encoded protein in a variety of ways which will lead to increases in the production of desired protein. The cited references show that the art was clearly aware that a combination of changes in codon preference and removal of sequences detrimental to transcription and/or translation in either the wild type gene or the codon optimized version can be used to accomplish this goal. While each of the cited references used a different combination of types of modifications, the art clearly teaches all of the claimed modifications encompassed in applicants claims (i.e., mammalian codon optimization, removal of transcription factor binding sequences, removal of splice sites, removal of potential promoters, and removal of polyadenylation sites) and clearly teaches combinations of them with one or more of the others.

Applicants argue that while there is a general teaching in the combination of cited documents to alter codons and/or remove certain undesired sequences in a selected sequence, none of the cited documents teaches or suggests that codon alterations, to prepare a sequence with codons employed more frequently in an evolutionarily divergent organism optionally in conjunction with

Art Unit: 1652

removal of restriction enzyme sites, ATTTA sequences, splice sites, polyA sites, A or T strings, CG dinucleotides in adjacent codons, prokaryotic promoters, inverted repeats and prokaryotic factor-independent RNA polymerase terminators, may create transcription factor binding sites and none of the cited documents discloses or suggests removal of transcription factor binding sites from a codon optimized gene. While it is true that none of the cited documents explicitly teach that codon replacements may create unwanted transcription factor binding sequences not present in the wild type sequence, Hey et al., Donnelly et al. and Pan et al. all show that the art recognized that codon modifications can **introduce** sequences which are unwanted within the synthetic gene, that additional codon modifications can decrease the introduction of those sequences and Sherf et al. clearly teach that the presence of transcription factor binding sequences within a reporter gene is an unwanted feature as it may interfere with the desired genetic neutrality of the reporter gene (see column 8). Furthermore, it is obvious on its face that anytime a gene sequence is altered that one necessarily creates new sequences which were not previously present and that merely by random chance some of these newly created sequences may be detrimental. It is even further obvious on its face that the more changes one makes, the

Art Unit: 1652

higher the chances that such a detrimental sequence will be introduced. Sherf et al. made only limited changes to codon selection and thus at least in his explicit teachings focused on the elimination of detrimental sequences present in the wild type sequence. However, the remaining art clearly would have motivated one of skill in the art to make more substantial changes in codon preference within the luciferase of Sherf et al. Furthermore the disclosures of Hey et al., Donnelly et al. and Pan et al. would have clearly led a skilled artisan to scan not only the wild type sequence for the unwanted transcription factor binding sites but also the codon optimized version thereof.

Applicants argue that to arrive at applicant's invention, one of skill in the art in possession of the cited documents would need to choose to identify specifically transcription factor binding sites, promoter sequences, splice sites, and polyA sites, as sequences to be removed by codon replacement although the references also teach removal of internal palindromic sequences, restriction endonuclease sites, glycosylation sites, ATTTA sequences, RNA polymerase termination signals, TA and CG doublets, blocks of G or C residues, inverted repeats, and long runs of purines. However, this is not persuasive as applicants claims are not drawn to any combination

---



in particular and do not exclude removal of other detrimental sequences in addition to those specifically recited in the claims and the art clearly teaches several combinations of these.

Applicants argue that none of the cited documents discloses or suggests the use of software to identify particular regulatory sites, such as mammalian transcription factor binding sequences, in a database of transcription factor binding sequences. However, this is not persuasive as most of applicants claims do not even mention the use of software to identify sites to be removed. Furthermore, even for those claims that do mention this, it is noted that the claims recite products not processes. Patentability of a product recited in product-by-process format is determined by the characteristics of the product itself not by the recited method. A nucleic acid in which the sites to be removed were identified by an undefined computer program would not differ in any respect from a nucleic acid in which the sites to be removed were identified by any other method.

Applicants finally argue that one of ordinary skill in the art in possession of the cited art would have no reasonable expectation that any particular set of changes would improve activity in a gene that is to be expressed in a highly

Art Unit: 1652

evolutionarily distinct cell. This is not persuasive because the art clearly provide an expectation that codon optimization and the elimination of a variety of types of sequences which are detrimental to transcription and/or translation will improve the expression of a gene in a heterologous host. While it is acknowledged that one cannot be certain that the modifications will not have unexpected consequences, applicants are reminded that obviousness does not require an absolute certainty of success but only a reasonable expectation thereof.

Claims 18, 47, 71, 74, 76-78, 80, 82-85, 87, 88 and 96 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sherf et al. (US Patent 5,670,356) in view of Zolotukhin et al. (US Patent 5,874,304), Donnelly et al. (WO 97/47358), Pan et al., Cornelissen et al. (US Patent 5,952,547), and Hey et al. (US Patent 6,169,232) as applied to claims 1, 3-6, 9, 11, 12, 15, 20, 21, 24-39, 41-45, 60, 67, 69, 70, 81, 86 and 90-95 above, and further in of Wood et al. (WO 99/14336). The rejection is explained in the previous Office Action.

Applicant has not presented any arguments specifically traversing this rejection but instead relies upon the traversal discussed above. Therefore, this rejection is maintained for the reasons presented above.

Art Unit: 1652

The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory obviousness-type double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., In re Berg, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); In re Goodman, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); In re Longi, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); In re Van Ornum, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); In re Vogel, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and In re Thorington, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent either is shown to be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement.

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

Claims 91, 93 and 94 are provisionally rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claims 1-50 and 58-60 of copending Application No. 10/314,827. Although the conflicting claims are not identical, they are not patentably distinct from each other. The rejection is explained in the previous Office Action.

Applicants argue that the claims in the present application are directed to synthetic nucleic acid molecules for

Art Unit: 1652

chloramphenicol acetyltransferase, Renilla luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase while the claims of 10/314,827 are directed to 10/314,827) are directed to synthetic nucleic acid molecules for a fluorescent polypeptide. However, it is noted that applicants have not amended claims 91, 93 and 94 of the instant application to synthetic nucleic acid molecules for chloramphenicol acetyltransferase, Renilla luciferase, beetle luciferase, beta-lactamase, beta-glucuronidase or beta-galactosidase. These claims recite synthetic nucleic acid molecules encoding any reporter polypeptide which clearly includes the fluorescent polypeptides of the copending application.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will

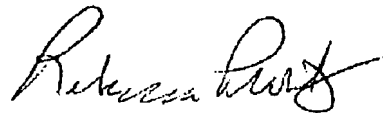
Art Unit: 1652

expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

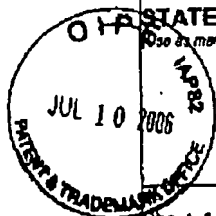
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Rebecca E. Prouty whose telephone number is 571-272-0937. The examiner can normally be reached on Tuesday-Friday from 8 AM to 5 PM. The examiner can also be reached on alternate Mondays

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ponnathapura Achutamurthy, can be reached at (571) 272-0928. The fax phone number for this Group is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Rebecca Prouty  
Primary Examiner  
Art Unit 1652



Substitute for form 1449A/PTO  
**INFORMATION DISCLOSURE  
 STATEMENT BY APPLICANT**  
 (Use as many sheets as necessary)

Complete if Known

**Application Number** 09/645,706  
**Filing Date** August 24, 2000  
**First Named Inventor** Wood, Keith  
**Group Art Unit** 1652  
**Examiner Name** Prouty, Rebecca

**Attorney Docket No:** 341.005US1

Sheet 1 of 10

**US PATENT DOCUMENTS**

Examiner Initial *	USP Document Number	Publication Date	Name of Patentee or Applicant of cited Document	Filing Date If Appropriate
RP -	US-2003/0157643A1	08/21/2003	Almond, B. D., et al.	12/09/2002
	US-2006/0127988A1	06/15/2006	Wood, K. V., et al.	12/22/2005
	US-4,246,340	01/20/1981	Lundin, A. T., et al.	05/30/1979
	US-4,412,001	10/25/1983	Baldwin, T. O., et al.	12/03/1981
	US-4,503,142	03/05/1985	Berman, M. L., et al.	06/25/1982
	US-4,581,335	04/08/1986	Baldwin, T. O., et al.	12/01/1982
	US-5,106,732	04/21/1992	Kondo, K., et al.	02/14/1990
	US-5,168,062	12/01/1992	Stinski, M. F.	09/10/1990
	US-5,196,524	03/23/1993	Gustafson, G. D., et al.	01/06/1989
	US-5,221,623	06/22/1993	Legocki, R. P., et al.	07/19/1989
	US-5,605,793	02/25/1997	Stemmer, W. P.	02/17/1994
	US-5,618,682	04/08/1997	Scheirer, W.	02/08/1994
	US-5,629,168	05/13/1997	Kricka, Larry J.	06/07/1995
	US-5,744,307	04/28/1998	Kuroda, N., et al.	09/12/1995
	US-5,814,471	09/29/1998	Wood, K. V.	04/09/1997
RP	US-6,602,677	08/05/2003	Wood, K. V., et al.	98/15/1999

**FOREIGN PATENT DOCUMENTS**

Examiner Initials *	Foreign Document No	Publication Date	Name of Patentee or Applicant of cited Document	T*
RP	EP-0337349A2	10/18/1989	Kajiyama, N., et al.	
	EP-0364707A1	04/25/1990	Tatsumi, H., et al.	
	EP-0437013A2	07/17/1991	Smith, D. F., et al.	
	EP-0449621A1	10/02/1991	Kajiyama, N.	
	EP-0524448A1	01/27/1993	Kajiyama, N., et al.	
	EP-0353464B1	10/20/1993	Tatsumi, H., et al.	
	JP-03-167288	07/19/1991	Inoue, S.	✓
	JP-07-067696 (w/ English Abstract)	03/14/1995	Yoshitami, M., et al.	
	JP-09-294600 (w/ English Abstract)	11/18/1997	Tatsumi, H., et al.	
	JP-10-87621	04/07/1998	Masako, M.	✓
	WO-90/01542A1	02/22/1990	Kazami, J., et al.	
	WO-91/16432A1	10/31/1991	Cornelissen, M., et al.	
	WO-92/15673A1	09/17/1992	Cornier, M. J., et al.	
	WO-95/18853A1	07/13/1995	Wood, K. V., et al.	
	WO-9726333A1	07/24/1997	Zolotukhin, S., et al.	
	WO-98/13487A1	04/02/1998	Stemmer, W. P., et al.	
RP	WO-98/46729A2	10/22/1998	Squirrell, D., et al.	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTO-1449)  
 \* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicants. Applicant's unique citation designation number (optional). Applicant is to place a check mark here if English language translation is attached.

Substitute for form 1449A/PTO

# **INFORMATION DISCLOSURE STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it contains a valid OMB control number.

PTO/BA/MA/18.01  
Approved for use through 10/31/2002. OMB 551-0031  
US Patent & Trademark Office is a participant in the CONFIDENTIALITY OF COMMERCIAL INFORMATION PROGRAM.

Complete if Known

Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 2 of 10

Attorney Docket No: 341.005US1

## **OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T*
RP		"Dual-Luciferase™ Reporter Assay System", (1998), 2 pgs.	
		"Luciferase Reporter Gene Technology", (1998), 4 pgs.	
		"Promega Technical Bulletin No. 161 - Luciferase Assay System With Reporter Lysis Buffer", (March, 1998), 9 pgs.	
		"Promega Technical Bulletin No. 101 - Luciferase Assay System", (March, 1998), 9 pgs.	
		"Promega Technical Manual - Dual-Luciferase™ Reporter Assay System", (February, 1999), 26 pgs.	
		"Promega Technical Manual - Steady-Glo™ Luciferase Assay System", (October, 1998), 19 pgs.	
		"Steady-Glo™ Luciferase Assay System", (1998), 2 pgs.	
		"TESS - Filtered String Search Page", <a href="http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=SEA-FR-QueryF">http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=SEA-FR-QueryF</a> , (June, 2000)	
		ALAM, J., et al., "Reporter Genes: Application to the Study of Mammalian Gene Transcription", <i>Analytical Biochemistry</i> , 188(2), (1990), 245-254	
		ANDREWS, E. M., et al., "Hierarchy of Polyadenylation Site Usage by Bovine Papillomavirus in Transformed Mouse Cells", <i>Journal of Virology</i> , 67(12), (December, 1993), 7705-7710	
		ARNOLD, F. H., "Directed Evolution: Creating Biocatalysts for the Future", <i>Chemical Engineering Science</i> , 51, (1996), 5091-5102	
		BACHMAIR, A., "In Vivo Half-Life of a Protein is a Function of its Amino Terminal Residue", <i>Science</i> , 234(4773), (1986), 179-186	
		BONIN, A. L., "Photinus pyralis Luciferase: Vectors that Contain a Modified luc Coding Sequence Allowing Convenient Transfer Into Other Systems", <i>Gene</i> , 141(1), (1994), 75-77	
		BOROVKOV, A. Y., et al., "XcmI-containing Vector for Direct Cloning of PCR Products", <i>BioTechniques</i> , 22(5), (May, 1997), 812-814	
		BOTHWELL, A. L., et al., "Heavy Chain Variable Region Contribution to the NP <sup>o</sup> Family of Antibodies: Somatic Mutation Evident in a γ2a Variable Region", <i>Cell</i> , 24(3), (June, 1981), 625-637	
RP		BOUTHORS, A.-T., "Site-Directed Mutagenesis of Residues 164, 170, 171, 179, 220, 237 and 242 in PER-1 β-lactamase Hydrolysing Expanded-Spectrum Cephalosporins", <i>Protein Engineering</i> , 12(4), (1999), 313-318	
		BOWIE, J. U., et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions", <i>Science</i> , 247(4948), (March 16, 1990), 1308-1310	
		BRANCHINI, B. R., et al., "Naphthyl- and Quinolylluciferin: Green and Red Light Emitting Protein Luciferin Analogues", <i>Photochemistry and Photobiology</i> , 49(5), (1989), 689-695 not submitted	
RP		BRONSTEIN, I. et al., "Chemiluminescent and Bioluminescent Reporter Gene Assays", <i>Analytical Biochemistry</i> , 219(2), (1994), 169-181	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTD-1449)

\* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. Applicant's unique citation designation number (optional) if Applicant is to place a check mark here if English language translation is attached

Substitute for form 1448A/PTO  
**INFORMATION DISCLOSURE  
 STATEMENT BY APPLICANT**  
 (Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Approved for use through 10/31/2002, OMB 631-0001  
 US Patent, Trademark Office is a DEPARTMENT OF COMMERCE

Complete if Known	
Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 3 of 10

Attorney Docket No: 341.005US1

**OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS**

Examiner Initials	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T <sup>1</sup>
RP		CADWELL, R. C., et al., "Randomization of Genes by PCR Mutagenesis", <u>PCR Methods and Applications</u> , 2, (1992), 28-33	
		CARSWELL, S., et al., "Efficiency of Utilization of the Simian Virus 40 Late Polyadenylation Site: Effects of Upstream Sequences", <u>Molecular and Cellular Biology</u> , 9(10), (1989), 4248-4258	
		CHEN, C.-Y. A., "Interplay of Two Functionally and Structurally Distinct Domains of the c-fos AU-Rich Element Specifies Its mRNA-Destabilizing Function", <u>Molecular and Cellular Biology</u> , 14(1), (January, 1994), 416-426	
		DE WET, J. R., "Cloning of Firefly Luciferase cDNA and the Expression of Active Luciferase in <i>Escherichia coli</i> ", <u>Proc. Natl. Acad. Sci. USA</u> , 82, (1985), 7870-7873	
RP		DE WET, J. R., "Firefly Luciferase Gene: Structure and Expression in Mammalian Cells", <u>Molecular and Cellular Biology</u> , 7(2), (1987), 725-737	
		DELUCA, M., et al., "Role of Sulfhydryl Groups in Firefly Luciferase", <u>Biochemistry</u> , 3(7), (1964), 935-939	
		DELUCA, M., et al., "The Hydrolase Properties of Firefly Luciferase", <u>Biochemical and Biophysical Research Communications</u> , 18(5-6), (1965), 836-842 not submitted	
RP		DELUCA, M., et al., "The Role of 1,N <sup>6</sup> -Ethenoadenosine Triphosphate and 1, N <sup>6</sup> -Ethenoadenosine", <u>Proc. Nat. Acad. Sci. USA</u> , 70(6), (1973), 1664-1666	
		DEMENTIEVA, E. I., "Physicochemical Properties of Recombinant <i>Luciola mingrelica</i> Luciferase and Its Mutant Forms", <u>Biochemistry</u> , 61 (1), (1996), 115-119	
		FARR, A., et al., "A Pitfall of Using a Second Plasmid to Determine Transfection Efficiency", <u>Nucleic Acids Research</u> , 20(4), (February 25, 1992), Page 920	
		FLEER, R., "High-Level Secretion of Correctly Processed Recombinant Human Interleukin-1 $\beta$ in <i>Kluyveromyces fragilis</i> ", <u>Gene</u> , 107(2), (1991), 285-295	
		FORD, S. R., et al., "Enhancement of Firefly Luciferase Activity by Cytidine Nucleotides", <u>Analytical Biochemistry</u> , 204(2), (1992), 283-291	
		FRAMPTON, J., et al., "Synergy Between the NF-E1 Erythroid-Specific Transcription Factor and the CACCC Factor in the Erythroid-Specific Promoter of the Human Porphobilinogen Deaminase Gene", <u>Molecular and Cellular Biology</u> , 10(7), (July, 1990), 3838-3842	
		FROMANT, M., et al., "Direct Random Mutagenesis of Gene-Sized DNA Fragments Using Polymerase Chain Reaction", <u>Analytical Biochemistry</u> , 224, (1995), 347-353	
		GLUZMAN, Y., "SV40-Transformed Cells Support the Replication of Early SV40 Mutants", <u>Cell</u> , 23(1), (January, 1981), 175-182	
RP		GOULD, S. J., "A Conserved Tripeptide Sorts Proteins to Peroxisomes", <u>The Journal of Cell Biology</u> , 108, (1989), 1657-1664	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTO-1448)

\* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 2001. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. \* Applicant's unique citation designation number (optional): Applicant is to place a check mark here if English language Translation is attached



Substitute for form 1449A/PTO <b>INFORMATION DISCLOSURE          STATEMENT BY APPLICANT</b> (Use as many sheets as necessary)	Complete if Known	
	Application Number	09/645,706
	Filing Date	August 24, 2000
	First Named Inventor	Wood, Keith
	Group Art Unit	1652
	Examiner Name	Prouty, Rebecca
Sheet 4 of 10	Attorney Docket No: 341.005US1	

OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS			
Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T
RP		GOULD, S. J., "Antibodies Directed Against the Peroxisomal Targeting Signal of Firefly Luciferase Recognize Multiple Mammalian Peroxisomal Proteins", <u>The Journal of Cell Biology</u> , 110, (1990), 27-34	
		GOULD, S. J., et al., "Firefly Luciferase as a Tool in Molecular and Cell Biology", <u>Analytical Biochemistry</u> , 175, (1988), 5-13	
		GOULD, S. J., "Identification and Characterization of a Peroxisomal Targeting Signal", <u>Dissertation Abstracts International</u> , Vol. 50/07-B, (1989), 2 pgs.	
		HANAHAN, D., "Chapter 4 - Techniques for Transformation of <i>E. coli</i> ", in: <u>DNA Cloning: A Practical Approach</u> , Volume 1, Chapter 6, Glover, D.W., (Editor), IRL Press, Oxford, (1985), 109-135	
		HASTINGS, J. W., "Biological Diversity, Chemical Mechanisms, and the Evolutionary Origins of Bioluminescent Systems", <u>Journal of Molecular Evolution</u> , 19(3/4), (1983), 309-321	
		HOFTE, H., et al., "Structural and Functional Analysis of a Cloned Delta Endotoxin of <i>Bacillus thuringiensis</i> Berliner 1715", <u>European Journal of Biochemistry</u> , 161, (1986), 273-280	
		HOLCIK, M., et al., "Four Highly Stable Eukaryotic mRNAs Assemble 3' Untranslated Region RNA-Protein Complexes Sharing cis and Trans Components", <u>Proc. Natl. Acad. Sci. USA</u> , 84, (March, 1987), 2410-2414	
		HSIEH, C.-J., "Nucleotide Sequence, Transcriptional Analysis, and Glucose Regulation of the Phenoxazinone Synthase Gene ( <i>phasA</i> ) from <i>Streptomyces antibioticus</i> ", <u>Journal of Bacteriology</u> , 177(20), (October, 1995), 5740-5747	
		JANOWSKI, M., "Ras Proteins and the ras-Related Signal Transduction Pathway", <u>Radiation and Environmental Biophysics</u> , 30(3), (1991), 185-189	
		JENSEN, P. R., et al., "The Sequence of Spacers Between the Consensus Sequences Modulates the Strength of Prokaryotic Products", <u>Applied and Environmental Microbiology</u> , 64(1), (January, 1998), 82-87	
		JOHNSON, L. R., "Role of the Transcription Factor Sox-2 in the Expression of the FGF-4 Gene in Embryonal Carcinoma Cells", <u>Molecular Reproduction and Development</u> , 50(4), (1998), 377-386	
		JONES, P. L., "Tumor Necrosis Factor Alpha and Interleukin-1 $\beta$ Regulate the Murine Manganese Superoxide Dismutase Gene Through a Complex Intronic Enhancer Involving C/EBP- $\beta$ and NF- $\kappa$ B", <u>Molecular and Cellular Biology</u> , 17(12), (1997), 6970-6981	
		KAJIYAMA, N., et al., "Enhancement of Thermostability of Firefly Luciferase from <i>Luciola lateralis</i> by a Single Amino Acid Substitution", <u>Bioscience, Biotechnology, and Biochemistry</u> , 58(6), (June 1994), 1170-1171	
	RP	KAJIYAMA, N., et al., "Isolation and Characterization of Mutants of Firefly Luciferase Which Produce Different Colors of Light", <u>Protein Engineering</u> , 4(6), (August 1991), 691-693	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute for form 1449AMPTO  
**INFORMATION DISCLOSURE  
 STATEMENT BY APPLICANT**  
 (Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

PTO/SBOM (10-01)  
 Approved for use through 10/01/2002. OMB 03-1-0001  
 US Patent & Trademark Office U.S. DEPARTMENT OF COMMERCE

Complete if Known	
Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca
Attorney Docket No: 341.005US1	

Sheet 5 of 10

**OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T
RP		KAJIYAMA, N, et al., "Thermostabilization of Firefly Luciferase by a Single Amino Acid Substitution at Position 217", <u>Biochemistry</u> , 32(50), (December 21, 1993), 13795-13799	
		KAO, R., et al., "Single Amino Acid Substitutions Affecting the Specificity of the Fungal Ribotoxin Mitogillin", <u>FEBS Letters</u> , 466(1), (January 21, 2001), 87-90	
		KAY, S. A., et al., "Video Imaging of Regulated Firefly Luciferase Activity in Transgenic Plants and <i>Drosophila</i> ", <u>Promega Notes Magazine</u> , 49, (1994), 7 pgs.	
		KUPRASH, D V., "Conserved $\kappa$ B Element Located Downstream of the Tumor Necrosis Factor $\alpha$ Gene: Distinct NF- $\kappa$ B Binding Pattern and Enhancer Activity in LPS Activated Murine Macrophages", <u>Oncogene</u> , 11(1), (1995), 97-106	
		KUTUZOVA, G. D., et al., "Bioluminescence Color Variation and Kinetic Behavior Relationships Among Beetle Luciferases", <u>Bioluminescence and Chemiluminescence, Molecular Reporting with Photons</u> , J W Hastings et al., (Editors), John Wiley & Sons, Chichester, England, (1996), 248-252	
		LAMB, K A., "Effects of Differentiation on the Transcriptional Regulation of the FGF-4 Gene: Critical Roles Played by a Distal Enhancer", <u>Molecular Reproduction and Development</u> , 51(2), (1998), 218-224	
		LATHE, R., "Synthetic Oligonucleotide Probes Deduced From Amino Acid Sequence Data Theoretical and Practical Considerations", <u>Journal of Molecular Biology</u> , 183, (1985), 1-12	
		LEE, R. T., "Substrate-Binding Properties of Firefly Luciferase", <u>Archives of Biochemistry and Biophysics</u> , 141(1), (1970), 38-52	
		LEWIS, M. K., et al., "Efficient Site Directed <i>in vitro</i> Mutagenesis Using Ampicillin Selection", <u>Nucleic Acids Research</u> , 18(12), (1990), 3439-3443	
		LORENZ, W. W., et al., "Isolation and Expression of a cDNA Encoding <i>Renilla reniformis</i> Luciferase", <u>Proc. Nat. Acad. Sci. USA</u> , 88, (May, 1991), 4438-4442	
		LUCAS, M., et al., "Coelenterazine Is a Superoxide Anion-Sensitive Chemiluminescent Probe: Its Usefulness in the Assay of Respiratory Burst in Neutrophils", <u>Analytical Biochemistry</u> , 206(2), (November 1, 1992), 273-277	
		MANUKHOV, I. V., et al., "Cloning of the <i>Vibrio harveyi</i> luxA and luxB Genes and the Expression of Bioluminescence in <i>Escherichia coli</i> and <i>Bacillus subtilis</i> ", <u>Russian Biotechnology</u> , 1, (1996), 1-6	
		MARANVILLE, E., et al., "Assessment of Amino-Acid Substitutions at Tryptophan 16 in $\sigma$ -galactosidase", <u>European Journal of Biochemistry</u> , 267(5), (2000), 1495-1501	
		MATSUMURA, I., et al., "Directed Evolution of the Surface Chemistry of the Reporter Enzyme $\beta$ -glucuronidase", <u>Nature Biotechnology</u> , 17, (July, 1999), 696-701	
RP		MATTHEWS, J., "Purification and Properties of <i>Renilla reniformis</i> Luciferase.", <u>Biochemistry</u> , 16(1), (January 11, 1977), 85-91	

EXAMINER /Rebecca Prouty/

DATE CONSIDERED 08/30/2006

Substitute Disclosure Summary Form (PTO-1449)  
 \* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 809. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. Applicant's unique citation designation number (optional). Applicant is to place a check mark here if English language translation is attached.

Substitute for form 1449A/PTO

**INFORMATION DISCLOSURE  
STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

PTO/USDA/10011  
Approved for use through 10/1/2002. OMB 0331-0201  
US Patent & Trademark Office, U.S. DEPARTMENT OF COMMERCE

Complete if Known

Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 6 of 10

Attorney Docket No: 341.005US1

**OTHER DOCUMENTS – NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T
RP		MCELROY, W. D., "Factors Influencing the Response of the Bioluminescent Reaction to Adenosine Triphosphate", <u>Archives of Biochemistry and Biophysics</u> , 22, (1949), 420-433	
		MCELROY, W. D., et al., "Function of Adenosine Triphosphate in the Activation of Luciferin", <u>Archives of Biochemistry and Biophysics</u> , 64, (1956), 257-271	
		MCELROY, W. D., et al., "Mechanisms of Bioluminescent Reactions", in: <u>A Symposium on Light and Life</u> , McElroy, W. D., et al., Editors, (John Hopkins Press, 1961), 219-257	
		MCWHERTER, C. A., et al., "Scanning Alanine Mutagenesis and De-Peptidization of a <i>Candida albicans</i> Myristoyl-CoA: Protein N-Myristoyltransferase Octapeptide Substrate Reveals Three Elements Critical for Molecular Recognition", <u>Journal of Biological Chemistry</u> , 272(18), (1997), 11874-11880	
		MOUNT, S. M., "Genomic Sequence, Splicing, and Gene Annotation", <u>American Journal of Human Genetics</u> , 67(4), (2000), 788-792	
		MOYER, J. D., et al., "Nucleoside Triphosphate Specificity of Firefly Luciferase", <u>Analytical Biochemistry</u> , 131(1), (1983), 187-189	
		MURAKAMI, S., et al., "Bioluminescent Enzyme Immunoassay Using Thermostable Mutant Luciferase and Acetate Kinase as a Labelled Enzyme", <u>Analytica Chimica</u> , 361, (1998), 19-26	
		MURRAY, I. A., et al., "Steroid Recognition by Chloramphenicol Acetyltransferase: Engineering and Structural Analysis of a High Affinity Fusidic Acid Binding Site", <u>Journal of Molecular Biology</u> , 254 (1995), 993-1005	
		NIBU, Y., "A Cell Type-Dependent Enhancer Core Element is Located in Exon 5 of the Human Angiotensinogen Gene", <u>Biochemical Biophysical Research Communications</u> , 205(2), (1994), 1102-1108	
		OW, D. W., "Transient and Stable Expression of the Firefly Luciferase Gene in Plant Cells and Transgenic Plants", <u>Science</u> , 234(4778), (November 14, 1986), 856-859	
		PETIT, T., et al., "A Mutation Ser <sup>213</sup> /Asn in the Hexokinase 1 from <i>Schizosaccharomyces pombe</i> Increases Its Affinity for Glucoses", <u>Biochemical and Biophysical Research Communications</u> , 251(3), (October 29, 1998), 714-719	
		PINTO, M., et al., "Denaturation of Proteins During Heat Shock", <u>The Journal of Biological Chemistry</u> , 266(21), (1991), 13941-13946	
RP		PURDY, D., et al., "Heterologous Gene Expression in <i>Campylobacter coli</i> : The Use of Bacterial Luciferase in a Promoter Probe Vector", <u>FEMS Microbiology Letters</u> , 111(2-3), (August 1, 1993), 233-237	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTO-1449)

\* EXAMINER: initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. Applicant's unique citation designation number (optional) Applicant is to place a check mark here if English language Translation is attached

**INFORMATION DISCLOSURE  
STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

PTO/USPOA(10-01)  
Approved for use through 10/31/2002, OMB 031-0031  
US Patent & Trademark Office is a DEPARTMENT OF COMMERCE

Complete if Known

Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 7 of 10

Attorney Docket No: 341.005US1

**OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T*
RP		REESE, M G., "Large Scale Sequencing Specific Neural Networks for Promoter and Splice Site Recognition", <u>Biocomputing: Proceedings of the 1996 Pacific Symposium</u> , Hunter, L., et al., Editors, World Publishing Co., Singapore (Abstract Only), (1996), 1 pg.	
		REESE, M G., et al., "New Neural Network Algorithms for Improved Eukaryotic Promoter Site Recognition", <u>The Seventh International Genome Sequencing and Analysis Conference</u> , Hilton Head Island, South Carolina, (Abstract Only), (1995), 1 pg.	
		RHODES, W. C., "The Synthesis and Function of Luciferyl-adenylate and Oxyluciferyl-adenylate", <u>J. Biol. Chem.</u> , 233(6), (1958), 1528-1537	
		RIGGS, K. J., et al., "Common Factor 1 Is a Transcriptional Activator Which Binds in the c-myc Promoter, the Skeletal $\alpha$ -Actin Promoter, and the Immunoglobulin Heavy-Chain Enhancer", <u>Molecular and Cellular Biology</u> , 11(3), (March, 1991), 1765-1769	
		ROMMENS, J. M., et al., "cAMP-inducible Chloride Conductance in Mouse Fibroblast Lines Stably Expressing the Human Cystic Fibrosis Transmembrane Conductance Regulator", <u>Proc. Natl. Acad. Sci.</u> , 88, (1991), 7500-7504	
		ROSENDAHL, M. S., et al., "Dimensional Probing of the ATP Binding Site on Firefly Luciferase", <u>Photochemistry and Photobiology</u> , 35(6), (1982), 857-861	
		SAIKI, R. K., et al., "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase", <u>Science</u> , 239(4839), (January 29, 1988), 487-491	
		SAISANIT, S., "A Novel Enhancer, the Pro-B Enhancer, Regulates Id1 Gene Expression in Progenitor B Cells", <u>Mol. Cell. Biol.</u> , 15(3), (1995), 1513-1521	
		SALA-NEWBY, G., "Engineering a Bioluminescent Indicator for Cyclic AMP-Dependent Protein Kinase", <u>Biochemical Journal</u> , 279(Part 3), (November 1991), 727-732	
		SALA-NEWBY, G. B., et al., "Expression of Recombinant Firefly Luciferase in Prokaryotic and Eukaryotic Cells", <u>Biochemical Society Transactions</u> , 20, (1992), Pg. 143S	
		SALA-NEWBY, G. B., "Stepwise Removal of the C-Terminal 12 Amino Acids of Firefly Luciferase Results in Graded Loss of Activity", <u>Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology</u> , 1208(1), (1994), 155-160	
		SCHMIDT, E. V., et al., "The Cytomegalovirus Enhancer: A Pan-Active Control Element in Transgenic Mice", <u>Molecular and Cellular Biology</u> , 10(8), (1990), 4406-4411	
		SCHUTTE, B., et al., "Optimized Conditions for Cloning PCR Products Into an XcmI T-Vector", <u>BioTechniques</u> , 22(1), (January, 1997), 40-42, 44	
RP		SELIGER, H. H., "The Colors of Firefly Bioluminescence: Enzyme Configuration and Species Specificity", <u>Proc. Natl. Acad. Sci., USA</u> , 52(1), (1964), 75-81	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

(Substitute Disclosure Statement Form (PTO-1449))

\* EXAMINER: Initial if reference considered, whether or not citation is in accordance with MPEP 900. Draw line through citation if not in accordance and not considered. Include copy of this form with next communication to applicant. Applicant's unique design designation number (optional) Applicant is to place a check mark here if English language Translation is attached

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it carries a valid OMB control number.

Substitute for form 1449A/PTO <b>INFORMATION DISCLOSURE          STATEMENT BY APPLICANT</b> (Use as many sheets as necessary)	Complete if Known	
	Application Number	09/645,706
	Filing Date	August 24, 2000
	First Named Inventor	Wood, Keith
	Group Art Unit	1652
	Examiner Name	Prouty, Rebecca
Sheet 8 of 10	Attorney Docket No: 341.005US1	

OTHER DOCUMENTS -- NON PATENT LITERATURE DOCUMENTS			
Examiner Initials*	Cite No. <sup>1</sup>	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T <sup>2</sup>
RP		SENAPATHY, P., et al., "[16] Splice Junctions, Branch Point Sites, and Exons: Sequence Statistics, Identification, and Applications to Genome Project", <u>Methods in Enzymology</u> , 183, (1990), 252-278	
		SEOL, J. H., et al., "Site-Directed Mutagenesis of the Cys Residues in ClpA, the ATPase Component of Protease Tl (SlpAP) in <i>Escherichia coli</i> ", <u>Biological Chemistry</u> , 378(10), (Abstract Only), (1997), 1205-1209	
		SHERF, B. A., et al., "Dual-Luciferase™ Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and <i>Renilla</i> Luciferase Assays", <u>Promega Notes Magazine</u> , Number 57, (1996), 7 pgs.	
		SIROT, D., et al., "A Complex Mutant of TEM-1 $\beta$ -Lactamase With Mutations Encountered in Both IRT-4 and Extended-Spectrum TEM-15, Produced by an <i>Escherichia coli</i> Clinical Isolate", <u>Antimicrobial Agents and Chemotherapy</u> , 41(6), (June, 1997), 1322-1325	
		SOMMER, J. M., "In Vivo Import of Firefly Luciferase into the Glycosomes of <i>Trypanosoma brucei</i> and Mutational Analysis of the C-Terminal Targeting Signal", <u>Molecular Biology of the Cell</u> , 3, (1992), 749-759	
		STAPLETON, P. D., "Construction and Characterization of Mutants of the TEM-1 $\beta$ -Lactamase Containing Amino Acid Substitutions Associated With Both Extended-Spectrum Resistance and Resistance to $\beta$ -Lactamase Inhibitors", <u>Antimicrobial Agents and Chemotherapy</u> , 43(8), (August, 1999), 1881-1887	
		STEMMER, W. P., et al., "DNA Shuffling by Random Fragmentation and Reassembly: <i>In vitro</i> Recombination for Molecular Evolution", <u>Proc. Natl. Acad. Sci. USA</u> , 91(22), (October 25, 1994), 10747-10751	
		STEWART, C. L., et al., "Expression of Retroviral Vectors in Transgenic Mice Obtain by Embryo Infection", <u>The EMBO Journal</u> , 8(2), (February, 1987), 383-388	
		STRAUSS, E. C., "In vivo Protein-DNA Interactions at Hypersensitive Site 3 of the Human $\beta$ -globin Locus Control Region", <u>Proc. Natl. Acad. Sci. USA</u> , 89, (July, 1992), 5809-5813	
		SZITTNER, R., et al., "Nucleotide Sequence, Expression, and Properties of Luciferase Coded by <i>lux</i> Genes from a Terrestrial Bacterium", <u>The Journal of Biological Chemistry</u> , 265(27), (September 25, 1990) 16581-16587	
		TARPEY, M. M., et al., "Chemiluminescent Detection of Oxidants in Vascular Tissue. Lucigenin But Not Coelenterazine Enhances Superoxide Formation", <u>Circulation Research</u> , 84(10), (May 28, 1999), 1203-1211	
		TERANISHI, K., "Coelenterazine Analogs as Chemiluminescent Probe for Superoxide Anion", <u>Analytical Biochemistry</u> , 249(1), (June 15, 1997), 37-43	
RP		VOLADRI, R. K. R., "Structure-Function Relationships Among Wild-Type Variants of <i>Staphylococcus aureus</i> $\beta$ -Lactamase: Importance of Amino Acids 128 and 216", <u>Journal of Bacteriology</u> , 178(24), (December, 1996), 7248-7253	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTO-1449)

\* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 600. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. \* Applicant's unique citation designation number (optional) \* Applicant is to place a check mark here if English language Translation is attached

**INFORMATION DISCLOSURE  
STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

PTO/SB/09A(10-01)  
Approved for use through 10/16/2007. OMB 05-1-003  
US Patent & Trademark Office U.S. DEPARTMENT OF COMMERCE

Complete if Known

Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 9 of 10

Attorney Docket No: 341.005US1

**OTHER DOCUMENTS – NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T*
RP		WADA, K.-N., et al., "Codon Usage Tabulated From the GenBank Genetic Sequence Data", <u>Nucleic Acids Research</u> , 20(Suppl.), (1992), 2111-2118	
		WAGNER, E. F., et al., "Transfer of Genes Into Embryonal Carcinoma Cells by Retrovirus Infection: Efficient Expression From an Internal Promoter", <u>The EMBO Journal</u> , 4(3), (March, 1985), 663-666	
		WALKER, D. E., et al., "An Aspartic Acid at Amino Acid 108 Is Required To Rescue Infectious Virus After Transfection of a Poliovirus cDNA Containing a CGDD but Not SGDD Amino Acid Motif in 3D <sup>pol</sup> ", <u>Journal of Virology</u> , 69(12), (1995), 8173-8177	
		WHITE, P. J., et al., "Generation and Characterisation of a Thermostable Mutant of Luciferase from Photinus Pyralis", In: <u>Bioluminescence and Chemiluminescence, Fundamentals and Applied Aspects, Proceedings of the 8th International Symposium on Bioluminescence and Chemiluminescence</u> , John Wiley & Sons, Cambridge, (September 1994), 419-422	
		WHITE, P. J., "Improved Thermostability of the North American Firefly Luciferase: Saturation Mutagenesis at Position 354", <u>Biochemical Journal</u> , 319 (Part 2), (October 15, 1996), 343-350	
		WOOD, K. V., "Bioluminescent Click Beetles Revisited", <u>Journal of Bioluminescence and Chemiluminescence</u> , 4(1), (July, 1989), 31-39	
		WOOD, K. V., "Complementary DNA Coding Click Beetle luciferases Can Elicit Bioluminescence of Different Colors", <u>Science</u> , 244(4905), (May 12, 1989), 700-702	
		WOOD, KEITH V., "Introduction to Beetle Luciferases and Their Applications", <u>Journal of Bioluminescence and Chemiluminescence</u> , 4, (July, 1989), 289-301	
		WOOD, K. V., "Luc Genes: Introduction of Colour Into Bioluminescence Assays", <u>Journal of Bioluminescence and Chemiluminescence</u> , 5, (April, 1990), 107-114	
		WOOD, K. V., "Photographic Detection of Luminescence in <i>Escherichia coli</i> Containing the Gene for Firefly Luciferase", <u>Analytical Biochemistry</u> , 161(2), (March 1987), 501-507	
		WOOD, K. V., et al., "Synthesis of Active Firefly Luciferase by In Vitro Translation of RNA Obtained From Adult Lanterns", <u>Biochemical and Biophysical Research Communications</u> , 124(2), (October 30, 1984), 592-596	
		WOOD, K. V., "The Chemical Mechanism and Evolutionary Development of Beetle Bioluminescence", <u>Photochemistry and Photobiology</u> , 62 (4), (1995), 662-673	
RP		YANAI, K., et al., "A cis-acting DNA Element Located Between TATA box and Transcription Initiation site Is Critical in Response to Regulatory Sequences in Human Angiotensinogen Gene", <u>Journal of Biological Chemistry</u> , 271(27), (1996), 15981-15986	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

Substitute Disclosure Statement Form (PTO-1449)

\* EXAMINER: Initial if reference considered, whether or not station is in conformance with MPEP 903. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. Applicant's unique design designation number (ppubnd) : Applicant is to place a check mark here if English language Translation is attached

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Substitute for form 1449A/PTO  
**INFORMATION DISCLOSURE  
 STATEMENT BY APPLICANT**  
 (Use as many sheets as necessary)

Complete if Known

Application Number	09/645,706
Filing Date	August 24, 2000
First Named Inventor	Wood, Keith
Group Art Unit	1652
Examiner Name	Prouty, Rebecca

Sheet 10 of 10

Attorney Docket No: 341.005US1

**OTHER DOCUMENTS – NON PATENT LITERATURE DOCUMENTS**

Examiner Initials*	Cite No. <sup>1</sup>	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume/issue number(s), publisher, city and/or country where published.	T <sup>2</sup>
RP		YE, L., "Cloning and Sequencing of a cDNA for Firefly Luciferase from <i>Photuris pennsylvanica</i> ", <i>Biochimica et Biophysica Acta</i> , 1339 (1), (April 25, 1997), 39-52	
RP		ZHANG, J.-H., et al., "Directed Evolution of a Fucosidase From a Galactosidase by DNA Shuffling and Screening", <i>Proc. Natl. Acad. Sci. USA</i> , 94(9), (April 29, 1997), 4504-4509	
RP		ZHAO, H., et al., "Functional and Nonfunctional Mutations Distinguished by Random Recombination of Homologous Genes", <i>Proc. Natl. Acad. Sci. USA</i> , 94(15), (July 22, 1997), 7997-8000	

EXAMINER

/Rebecca Prouty/

DATE CONSIDERED

08/30/2006

(Substitute Disclosure Statement Form (PTO-1449))

\* EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. \* Applicant's unique citation designation number (optional): Applicant is to place a check mark here if English language Translation is attached

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Keith V. Wood et al.	Examiner:	Rebecca E. Prouty
Serial No.:	09/645706	Group Art Unit:	1652
Filed:	August 24, 2000	Docket No.:	341.005US1
Title:	SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION		

---

DECLARATION UNDER 37 C.F.R. § 1.132

Commissioner for Patents  
Washington, D.C. 20231

Sir:

I, Monika Wood, M.S., declare and say as follows:

1. I am one of the named co-inventors of the claims in the above-identified application. I make this Declaration in support of the patentability of the claims of the above-identified application.
2. Sherf et al. (U.S. Patent No. 5,670,356) disclose that a firefly luciferase (*luc*) gene was modified using mammalian codon replacement to remove 3 internal palindromic sequences, 5 restriction endonuclease sites, 4 glycosylation sites, and 6 transcription factor binding sites, yielding *luc+*. On June 14<sup>th</sup> 2006, using publicly available software and a database of transcription factor binding sites (see attached details on the specific software, search parameters, and database release used), comparable to those employed in the above-referenced application, potential mammalian transcription factor binding sites were identified in the *luc+* gene. I found that the *luc+* gene contains over 150 potential mammalian transcription factor binding sites.
3. Thus, mammalian transcription factor binding sites in a particular nucleic acid sequence can be identified and enumerated.
4. I further declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title



DECLARATION UNDER 37 CFR § 1.132

Serial Number: 09/645706

Filing Date: August 24, 2000

Title: SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION

Page 2

Dkt.: 341.005US1

18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Dated: June-19-2006

By: M Wood  
Monika Wood

---

**Codon usage tabulated from the GenBank Genetic Sequence Data**

---

Shin-ichi Aota, Takashi Gojobori, Fumie Ishibashi, Takeo Maruyama\* and Toshimichi Ikemura

---

National Institute of Genetics, Mishima, 411, Japan

---

In 1980 and 1981, Grantham and his colleague (1,2) reported the codon usages in a total of 161 protein genes in this journal, and in 1986 we reported those in 1638 genes (3). In the present work, the codon usages in 3681 genes are analyzed using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 50.0, 1987)(4). In selecting protein coding sequences we relied on the FEATURES tables of the GenBank Database, and only complete genes, starting with an initiation codon and ending with one of stop codons, were used in the analysis (see ref. 3 for details). Table 1 lists the codon use in each of the 3681 genes. The LOCUS names given in the GenBank were used for designating individual genes, and the SHORT DIRECTORY of the GenBank is presented after the table for defining each LOCUS name. In the GenBank, a group of consecutive genes whose entire region had been sequenced were registered under one LOCUS name. To distinguish the different genes belonging to a single LOCUS, symbol # followed by a number is added after the LOCUS name; the numbers represent the order of the peptides registered in the FEATURES of the GenBank. Thus the numbering system differs from the previous one (3) in the cases where incomplete peptides are registered in the FEATURES. When introns of a gene have not been completely sequenced, some of its exons are registered in separate entries (LOCUS) in the GenBank. These exons belonging to the same gene but having different LOCUS names were combined, and the LOCUS name of the last exon followed by symbol \* was given to the gene thus combined. The order of the codons in the table is the same as the previous compilation (1-3). The amino acids based on the "universal genetic code" are specified using three letter abbreviations, except for the pages listing organella genes.

To reveal the characteristics of the codon use of individual organisms, as well as viruses and organella, the frequency (per one thousand) of codon use in each organism for which more than 20 genes are available in Table 1

was calculated by adding for each codon (Table 2). The number of genes summed for each organisms is given in the row designated as No. GENES, and the total codon number thus summed is given at the bottom row. Since the codon usage of each organism thus summed has been expressed in frequency per one thousand in Table 2, it is easy to compare the codon-choice patterns among different organisms. In the previous work (3), we noted that the resultant codon-choice patterns among the vertebrates, or at least among the mammals, are very similar, although the codon-choices in individual genes of one mammal are often very different with each other (e.g., see ref. 5). We also mentioned that the codon-choice pattern, that are roughly common among the mammals, does not depend on the choice of genes; i.e., when the codon frequencies for ten or more genes with varying functions are summed up for each mammal, they usually result in a very similar pattern regardless of the genes compiled (6). Tables 1 and 2 confirmed the previous notion.

### ACKNOWLEDGMENTS

The authors are grateful to Mr. A. Motohashi (FACOM-HITAC Corp.) for technical assistance in computer analysis. This work was supported by grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan.

\* It is with great sorrow that we have to inform you of Dr. Maruyama's sudden death. He died of a heart attack on December 11 in 1987. Thus, this paper became his posthumous work. We express our condolences to his family.

### REFERENCES

1. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Paŕe, A. (1980) Nucl. Acids Res. 8, r49-62.
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-74.
3. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) Nucl. Acids Res. 14, r151-197.
4. Bilofsky, H.S., Burks, C., Fickett, J.W., Goad, W.B., Lewitter, F.L., Rindone, W.P., Swindell, C.D. and Tung, C.S. (1986) Nucl. Acids Res. 14, 1-4.
5. Aota, S. and Ikemura, T. (1986) Nucl. Acids Res. 14, 6345-6355 & 8702.
6. Ikemura, T. (1985) Mol. Biol. Evol. 2, 13-24.

TABLE 1.  
/ PRIMATES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	148
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----

2	8	16	24	32	40	48	56	64	72	80	88	96	104	112	120	128	136	144	152	160	168	176	184	192	200	208	216	224	232	240	248	256	264	272	280	288	296	304	312	320	328	336	344	352	360	368	376	384	392	400	408	416	424	432	440	448	456	464	472	480	488	496	504	512	520	528	536	544	552	560	568	576	584	592	600	608	616	624	632	640	648	656	664	672	680	688	696	704	712	720	728	736	744	752	760	768	776	784	792	800	808	816	824	832	840	848	856	864	872	880	888	896	904	912	920	928	936	944	952	960	968	976	984	992	1000	1008	1016	1024	1032	1040	1048	1056	1064	1072	1080	1088	1096	1104	1112	1120	1128	1136	1144	1152	1160	1168	1176	1184	1192	1200	1208	1216	1224	1232	1240	1248	1256	1264	1272	1280	1288	1296	1304	1312	1320	1328	1336	1344	1352	1360	1368	1376	1384	1392	1400	1408	1416	1424	1432	1440	1448	1456	1464	1472	1480	1488	1496	1504	1512	1520	1528	1536	1544	1552	1560	1568	1576	1584	1592	1600	1608	1616	1624	1632	1640	1648	1656	1664	1672	1680	1688	1696	1704	1712	1720	1728	1736	1744	1752	1760	1768	1776	1784	1792	1800	1808	1816	1824	1832	1840	1848	1856	1864	1872	1880	1888	1896	1904	1912	1920	1928	1936	1944	1952	1960	1968	1976	1984	1992	2000	2008	2016	2024	2032	2040	2048	2056	2064	2072	2080	2088	2096	2104	2112	2120	2128	2136	2144	2152	2160	2168	2176	2184	2192	2200	2208	2216	2224	2232	2240	2248	2256	2264	2272	2280	2288	2296	2304	2312	2320	2328	2336	2344	2352	2360	2368	2376	2384	2392	2400	2408	2416	2424	2432	2440	2448	2456	2464	2472	2480	2488	2496	2504	2512	2520	2528	2536	2544	2552	2560	2568	2576	2584	2592	2600	2608	2616	2624	2632	2640	2648	2656	2664	2672	2680	2688	2696	2704	2712	2720	2728	2736	2744	2752	2760	2768	2776	2784	2792	2800	2808	2816	2824	2832	2840	2848	2856	2864	2872	2880	2888	2896	2904	2912	2920	2928	2936	2944	2952	2960	2968	2976	2984	2992	3000	3008	3016	3024	3032	3040	3048	3056	3064	3072	3080	3088	3096	3104	3112	3120	3128	3136	3144	3152	3160	3168	3176	3184	3192	3200	3208	3216	3224	3232	3240	3248	3256	3264	3272	3280	3288	3296	3304	3312	3320	3328	3336	3344	3352	3360	3368	3376	3384	3392	3400	3408	3416	3424	3432	3440	3448	3456	3464	3472	3480	3488	3496	3504	3512	3520	3528	3536	3544	3552	3560	3568	3576	3584	3592	3600	3608	3616	3624	3632	3640	3648	3656	3664	3672	3680	3688	3696	3704	3712	3720	3728	3736	3744	3752
---	---	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

[illegible]

181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

5	1000	9000	8000	7000	6000	5000	4000	3000	2000	1000	0000	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	12000	13000	14000	15000	16000	17000	18000	19000	20000	21000	22000	23000	24000	25000	26000	27000	28000	29000	30000	31000	32000	33000	34000	35000	36000	37000	38000	39000	40000	41000	42000	43000	44000	45000	46000	47000	48000	49000	50000	51000	52000	53000	54000	55000	56000	57000	58000	59000	60000	61000	62000	63000	64000	65000	66000	67000	68000	69000	70000	71000	72000	73000	74000	75000	76000	77000	78000	79000	80000	81000	82000	83000	84000	85000	86000	87000	88000	89000	90000	91000	92000	93000	94000	95000	96000	97000	98000	99000	100000	101000	102000	103000	104000	105000	106000	107000	108000	109000	110000	111000	112000	113000	114000	115000	116000	117000	118000	119000	120000	121000	122000	123000	124000	125000	126000	127000	128000	129000	130000	131000	132000	133000	134000	135000	136000	137000	138000	139000	140000	141000	142000	143000	144000	145000	146000	147000	148000	149000	150000	151000	152000	153000	154000	155000	156000	157000	158000	159000	160000	161000	162000	163000	164000	165000	166000	167000	168000	169000	170000	171000	172000	173000	174000	175000	176000	177000	178000	179000	180000	181000	182000	183000	184000	185000	186000	187000	188000	189000	190000	191000	192000	193000	194000	195000	196000	197000	198000	199000	200000	201000	202000	203000	204000	205000	206000	207000	208000	209000	210000	211000	212000	213000	214000	215000	216000	217000	218000	219000	220000	221000	222000	223000	224000	225000	226000	227000	228000	229000	230000	231000	232000	233000	234000	235000	236000	237000	238000	239000	240000	241000	242000	243000	244000	245000	246000	247000	248000	249000	250000	251000	252000	253000	254000	255000	256000	257000	258000	259000	260000	261000	262000	263000	264000	265000	266000	267000	268000	269000	270000	271000	272000	273000	274000	275000	276000	277000	278000	279000	280000	281000	282000	283000	284000	285000	286000	287000	288000	289000	290000	291000	292000	293000	294000	295000	296000	297000	298000	299000	300000	301000	302000	303000	304000	305000	306000	307000	308000	309000	310000	311000	312000	313000	314000	315000	316000	317000	318000	319000	320000	321000	322000	323000	324000	325000	326000	327000	328000	329000	330000	331000	332000	333000	334000	335000	336000	337000	338000	339000	340000	341000	342000	343000	344000	345000	346000	347000	348000	349000	350000	351000	352000	353000	354000	355000	356000	357000	358000	359000	360000	361000	362000	363000	364000	365000	366000	367000	368000	369000	370000	371000	372000</
---	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	----------



[illegible]

[illegible]

	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
--	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 336

[illegible]

[illegible]



[illegible]

## 22

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

## / BACTERIA &amp; PLASMIDS

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

[illegible]

[illegible]

[illegible]



[illegible]



[illegible]

[illegible]

[illegible]

12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75																																						

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87													

[illegible]

[illegible]

ACT2 ACT3 ACT4 ACT5 ACT6

[illegible]

23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	15
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	----

2	4	5	2	144	141	142	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537	1538	1539	1540	1541	1542	1543	1544	1545	1546	1547	
---	---	---	---	-----	-----	-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



[illegible]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87													



[illegible]

161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

[illegible]

[illegible]

[illegible]



[illegible]

## / BACTERIOPHAGES

[illegible]

120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89											

[illegible]

[illegible]

67

001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023	024	025	026	027	028	029	030	031	032
033	034	035	036	037	038	039	040	041	042	043	044	045	046	047	048	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063	064
065	066	067	068	069	070	071	072	073	074	075	076	077	078	079	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095	096
097	098	099	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192
193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224
225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256
257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288
289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352
353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384
385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416
417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448
449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503									

“

69	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

710	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
AAA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
BBB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CCC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DDD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
EEE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81																			



[illegible]

[illegible]

[illegible]

[illegible]

MS. GENRE	407	31	184	230	49	25	71	23	44	21	153	53	407	28	22	29	33	37	59	61	34	58	54	45
ABO CCG	5.0	4.0	4.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	
ABO CCG	10.6	17.1	11.3	12.4	10.6	14.7	11.6	10.6	14.5	1.7	9.0	23.6	21.8	16.8	32.4	1.7	39.4	15.2	15.9	6.1	14.6	2.0	4.4	
ABO CCG	4.7	9.7	9.3	10.6	10.7	7.2	4.4	5.6	3.6	1.0	3.1	3.6	6.7	12.3	10.9	2.2	34.6	14.4	10.3	1.5	3.0	1.1	4.4	
ABO CCG	4.4	6.5	4.4	2.9	3.6	7.9	12.6	7.5	4.6	10.7	28.3	18.9	10.5	7.9	0.5	5.4	5.5	16.9	22.2	22.7	16.1	18.4		
ABO CCG	10.0	10.5	11.6	13.0	9.9	7.6	7.0	10.2	3.1	3.6	25.3	12.2	1.4	1.4	8.9	7.0	41.4	1.5	7.1	9.4	6.7	5.5	16.0	
ABO CCG	11.1	7.1	10.6	12.1	10.0	11.0	10.3	7.8	5.1	13.3	1.1	2.4	0.9	3.1	5.5	9.7	13.7	4.5	15.0	4.7	1.5	3.4	0.9	
ABO CCG	6.7	4.3	6.6	6.1	4.6	4.0	3.2	7.8	5.0	14.0	11.9	5.6	2.9	3.6	10.4	10.7	11.6	4.3	9.1	3.1	6.6	10.5	11.7	
ABO CCG	21.8	14.9	20.5	19.0	20.7	24.7	20.1	11.5	11.0	21.7	4.0	10.4	9.5	8.2	18.0	12.5	10.1	19.5	25.5	6.9	10.5	2.0	4.0	
ABO CCG	11.2	13.2	11.5	12.4	11.6	7.1	41.5	22.5	9.7	42.5	22.5	9.7	58.0	10.2	13.3	16.4	8.0	15.0	13.2	20.8	16.3	21.8	19.4	
ABO CCG	4.1	3.6	3.4	3.2	3.4	2.0	2.5	12.4	1.4	6.2	23.7	19.7	9.9	11.6	7.9	3.3	23.7	1.2	4.0	24.3	12.0	72.7	34.5	
ABO CCG	11.5	8.4	10.8	11.0	7.7	7.5	8.1	13.2	18.4	22.2	11.5	10.7	11.6	18.7	15.9	13.3	8.4	9.5	5.8	11.5	10.0	5.2	20.3	
ABO CCG	9.8	6.3	10.0	7.4	7.0	4.5	8.8	13.5	14.0	5.6	4.0	5.6	4.1	5.9	13.7	2.1	10.0	13.4	16.2	9.2	13.9	9.5		
ABO CCG	18.9	18.9	18.1	15.8	19.3	22.9	16.1	12.4	20.9	10.8	14.5	8.5	10.1	12.1	13.7	16.6	5.9	18.1	20.6	11.8	32.8	3.0	10.9	
ABO CCG	4.0	3.4	3.6	2.9	4.4	5.3	3.7	16.9	7.3	5.7	6.8	7.3	7.9	11.0	8.2	7.5	10.0	8.0	16.5	9.9	6.7	4.2	4.4	
ABO CCG	14.5	13.5	13.1	12.7	10.5	12.2	11.9	16.6	4.2	9.8	26.0	15.7	11.3	9.2	11.0	10.4	7.0	11.6	11.6	24.3	18.0	25.8	14.3	
ABO CCG	14.5	13.5	13.1	12.7	10.5	12.2	11.9	16.6	4.2	9.8	26.0	15.7	11.3	9.2	11.0	10.4	7.0	11.6	11.6	24.3	18.0	25.8	14.3	
ABO CCG	8.4	13.2	11.5	12.4	10.5	12.2	11.9	16.6	4.2	9.8	26.0	15.7	11.3	9.2	11.0	10.4	7.0	11.6	11.6	24.3	18.0	25.8	14.3	
ABO CCG	14.5	13.5	13.1	12.7	10.5	12.2	11.9	16.6	4.2	9.8	26.0	15.7	11.3	9.2	11.0	10.4	7.0	11.6	11.6	24.3	18.0	25.8	14.3	
ABO CCG	14.5	13.5	13.1	12.7	10.5	12.2	11.9	16.6	4.2	9.8	26.0	15.7												

r392

[illegible]

1394









YSC2P01 YEAST (S.CEREVISIAE) BSL1 GENE ENCODING INFORMATION REGULATORY GENE, 10256P  
YSC2P02 YEAST (S.CEREVISIAE) BWP1 GENE ENCODING A CARBON CATABOLITE 2657P  
YSC30000 YEAST GENE FOR MITOCHONDRIAL MANGANESE SUPEROXIDE DISMUTASE 2048P  
YSC8P72 YEAST (S.CEREVISIAE) BPT1 GENE ENCODING REGULATORY PROTEIN, 8025P  
YSC8P73 YEAST (S.CEREVISIAE) BGL1 TYPE-SPECIFIC BETA1S SITES GENE, 2028P  
YSC8P74 YEAST (S.CEREVISIAE) BPT2 GENE RECEPTOR FOR PHENOL GENE, 2028P  
YSC8P75 YEAST (S.CEREVISIAE) CEL1 TYPE-SPECIFIC STERILE SITES GENE, 2028P  
YSC8P76 YEAST (S.CEREVISIAE) REGULATORY GENE 8757, COMPLETE CDS, 2100P  
YSC8P77 YEAST (S.CEREVISIAE) SUC1 GENE ENCODING AN INVERTASE, COMPLETE CDS, 2028P  
YSC8P78 YEAST (S.CEREVISIAE) THB1 GENE ENCODING A HYDROLYTIC KINASE (TK), 2028P  
YSC8P79 YEAST (S.CEREVISIAE) ENCODING HYDROLYTIC KINASE (TK), 2028P  
YSC8P80 YEAST (S.CEREVISIAE) TOP2 GENE ENCODING TOPISOMERASE 2, COMPLETE CDS, 2028P  
YSC8P81 YEAST (S.CEREVISIAE) AUTONOMOUS REPLICATION SEQUENCE ARS1 AND THP1  
YSC8P82 YEAST (S.CEREVISIAE) THP1 GENE CODING FOR ANTHRANILATE SYNTHASE  
YSC8P83 YEAST (S.CEREVISIAE) THP1 GENE CODING FOR ANTHRANILATE SYNTHASE  
YSC8P84 YEAST (S.CEREVISIAE) BETA-TUBULIN GENE, 1425P  
YSC8P85 YEAST CLASS 8 TRANSCRIPTION TVT-1 7822P  
YSC8P86 YEAST (S.CEREVISIAE) URA3 GENE ENCODING OMPUTON 4-PHOSPHATE  
YSC8P87 YEAST (S.CEREVISIAE) YPS1 PROTEIN PROTO-CONJUGAL (YPS1-1) GENE, 1425P  
YSC8P88 YEAST (S.CEREVISIAE) EXT1 GENE ENCODING GENE STAI, 2753P  
YSC8P89 YEAST (S.CEREVISIAE) BETA-GLUCOSIDASE GENE, COMPLETE CDS, 4237P  
YSC8P90 YEAST (S.CARLSBERGENSES) ACTIN GENE, 1425P  
YSC8P91 YEAST (S.CARLSBERGENSES) MAL1 AND MAL2 GENE ENCODING MALTASE, 1425P  
YSC8P92 YEAST (S.CARLSBERGENSES) MEL1 (ALPHA-GALACTOSIDASE) GENE, 3703P  
YSC8P93 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P94 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P95 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P96 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P97 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P98 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P99 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P100 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P101 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P102 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P103 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P104 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P105 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P106 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P107 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P108 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P109 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P110 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P111 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P112 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P113 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P114 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P115 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P116 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P117 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P118 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P119 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P120 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P121 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P122 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P123 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P124 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P125 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P126 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P127 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P128 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P129 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P130 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P131 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P132 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P133 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P134 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P135 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P136 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P137 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P138 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P139 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P140 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P141 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P142 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P143 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P144 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P145 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P146 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P147 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P148 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P149 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P150 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P151 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P152 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P153 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P154 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P155 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P156 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P157 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P158 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P159 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P160 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P161 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P162 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P163 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P164 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P165 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P166 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P167 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P168 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P169 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P170 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P171 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P172 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P173 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P174 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P175 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC8P176 YEAST (S.CARLSBERGENSES) RIBONUCLEONIC PROTEIN, 1425P  
YSC



r400



[illegible]

PT4RE	BACTERIOPHAGE T2 (MUTANT) RIGHT END DNA, ENCODING TAIL FIBER 1447BP	YBOMTROF	YEAST (S.CEREVISIAE) MT 218 rRNA GENE INTRON R1 CONTAINING AN ORF.
PT4B07	BACTERIOPHAGE T4 BETA-GT GENE ENCODING BETA-GLUCOSYLTRANSFERASE	YBOMT05A	YEAST (S.CEREVISIAE) MITOCHONDRIAL SER-TRNA AND ATPASE PROTEOLIPID
PT4B0V	BACTERIOPHAGE T4 DEW GENE ENCODING ENDONUCLEASE V, COMPLETE CDS.	YBOMT0A1	YEAST (S.CEREVISIAE) MITOCHONDRIAL VAR1 GENE, 40.6 ALLELE, 1200BP
PT4DEXA	BACTERIOPHAGE T4 DEXA GENE, COMPLETE CDS, 1204BP AND 32 (COMPLETE CDS), 2200BP	YBOMT03	YEAST (T.OLIVIPATA) MITOCHONDRIAL DNA BETWEEN CYTOCHROME OXIDASE
PT4E	BACTERIOPHAGE T4 E GENE ENCODING LYKOTRIENE, COMPLETE CDS, 948BP	YBOMT05B	YEAST (S. POMBE) MITOCHONDRIAL COB GENE ENCODING APOCYTOCHROME B.
PT4FRD01	BACTERIOPHAGE T4 FRD AND TD GENES ENCODING DIPHOLATE REDUCTASE		
PT4FRD02	BACTERIOPHAGE T4 TD GENE ENCODING THYMIDYLATE SYNTHASE, EXON 2.		
PT4G1TRNA	BACTERIOPHAGE T4 GENES 1, 57A, 57B, 1P1, AND TRNA CLUSTERS, 5018BP		
PT4G2G03	BACTERIOPHAGE T4 GENES 22 (PARTIAL) AND 23 (COMPLETE CDS), 2200BP		
PT4G30	BACTERIOPHAGE T4 GENE 30 ENCODING DNA LIASE, 1247BP		
PT4G32	BACTERIOPHAGE T4 GENE 32 ENCODING 8S-DNA BINDING PROTEIN, 1240BP		
PT4G3037	BACTERIOPHAGE T4 GENES 36 AND 37 (COMPLETE); 36 AND 38 (PARTIAL)		
PT4G3043	BACTERIOPHAGE T4 GENES 34, ALPHA-GT, 47, 48, 49, 44, 62, REGA,		
PT4G3043C	BACTERIOPHAGE T4 GENES 33 (ENCODING RNA LIASE) AND ALC, COMPLETE		
PT4G3051	BACTERIOPHAGE T4 GENES 37, 38 AND 21, 75BP		
PT4P8ET	PHAGE T4 P8ET GENE FOR POLYNUCLEOTIDE KINASE (PNK), 2035P		
PT4UVBY	BACTERIOPHAGE T4 UVBY GENE INVOLVED IN THE RECOMBINATION PATHWAY.		
PT4UVBY90	BACTERIOPHAGE T4 GENES UVBY, 41, 616A, DAM, 99 AND SOC ENCODING		
PT5X00	BACTERIOPHAGE T5 D118 GENE FOR DNA LIASE, 1247BP		
PT5TRD0	BACTERIOPHAGE T5 GENE FOR TRNA-ASP (GUC), 754BP		
PT7	BACTERIOPHAGE T7, COMPLETE GENOME, 39438BP		
PT710	BACTERIOPHAGE T7 GENES 1.1, 1.2, 1.3 (POSITIONS 14.45%-18.15%), AND		
PX1	BACTERIOPHAGE PHX-X174, COMPLETE GENOME, 5398BP		
PZAMER	BACTERIOPHAGE PZA (FROM S.SUTLIS) MAJOR EARLY REGION, 5200BP		
SP1MP027	BACTERIOPHAGE SPO1 (S.SUTLIS) GENE 27, MIDDLE PROMOTER, AND GENE		
SP1S0	BACTERIOPHAGE SPO1 (S.SUTLIS) GENES 33 AND 34 ENCODING SIGMA		
SP1TF1	BACTERIOPHAGE SPO1 TRANSCRIPTION FACTOR 1 (TF1) GENE, COMPLETE CDS.		
SP2PCOL	BACTERIOPHAGE SPO2 (S.SUTLIS) GENE L ENCODING DNA POLYMERASE.		
SP2D	BACTERIOPHAGE SP2 (S.SUTLIS) D GENE ENCODING MINORITY 1033BP		
SP2CAN	BACTERIOPHAGE S-PALC (FROM S. AUREUS) SAK GENE CODING FOR 1377BP		
SP2MASE	BACTERIOPHAGE SPS (S.SUTLIS) DNA METHYLTRANSFERASE GENE, 2128BP		
----- ORGANELLA			
AMNCPP58A	ALYNDIDUS CHLOROPLAST HERBICIDE BINDING PROTEIN PSBA COMPLETE		
ASMTCTY2	ANIDULANS MT APOCYTOCHROME B (COB) GENE, EXON2, 1082BP		
ASMTMDA	ANIDULANS MT DNA BETWEEN H248 & H282 JUNCTIONS, 14440BP		
ASMTMDA	ANIDULANS MT ARG-TRNA, ASP-TRNA AND ATPASE (SUBUNIT 6) GENES.		
ASMTMURF	ANIDULANS MT UNIDENTIFIED READING FRAMES URF1 AND URF4, 2803BP		
BLVCFAT7B	BANLEY S C4E AND C5E TRNA GENES: 3810BP		
BOVMT	BOVINE MITOCHONDRION, COMPLETE GENOME, 16336BP		
CRCOPAT7B	CHLAMYDOMONAS REINHARDTII CHLOROPLAST ATP6 GENE ENCODING CP-1		
CRCOPRUB	C. REINHARDTII CP RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE SUBUNIT		
CRCOPRUBA	C. REINHARDTII CHLOROPLAST PSBA GENE, ENCODING 32 KD PROTEIN OF 417BP		
CRCOPRUBF	CHLAMYDOMONAS REINHARDTII MITOCHONDRIAL URU1 GENE, 2318BP		
DROMTM1	D. MELANOGASTER MT CYTOCHROME C OXIDASE SUBUNIT, ATPASE, 7 TRNA		
DRYMT00	DROSOPHILA MITOCHONDRIAL CYTOCHROME B AND FIVE TRNA GENES, 6442BP		
DRYMTY1	D. YAKUBA MT CYTOCHROME C OXIDASE II AND II, ATPASE, 4 TRNA GENES,		
DRYMTY02	DROSOPHILA YAKUBA MITOCHONDRIAL TRNA GENE CLUSTER: SER-TRNA, 1625BP		
EGRCPEFTU	EULENA GRACILIS CHLOROPLAST ELONGATION FACTOR TU GENE, 2066BP		
EGRCPP89	E. GRACILIS CHLOROPLAST PSBA GENE CODING FOR 32 KDA THYLAKOID 4804BP		
EGRCPP89A	EULENA GRACILIS CHLOROPLAST PSBA GENE ENCODING 32-KD HERBICIDE		
EGRCPP89C	EULENA GRACILIS CHLOROPLAST RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE		
EGRCPP87	E. GRACILIS CHLOROPLAST RIBOSOMAL PROTEIN 87 AND 819 GENES, 1356BP		
HUMMT	HUMAN MITOCHONDRION, COMPLETE GENOME, 16569BP		
LEKPMAX	LEISHMANIA TARENTOLAE KINETOPLAST MAJOR CIRCLE DNA COMPLETE 20991BP		
MPQPC0	MARCHANTIA POLYMEROPHIA (LIVERWORT) CHLOROPLAST COMPLETE GENOME,		
MZECATB	MAIZE (Z. MAIZE) CP COUPLING FACTOR COMPLEX (CF-1) BETA & EPSILON		
MZECPP61	MAIZE CHLOROPLAST PHOTOSYSTEM I P61A1 AND P61A2 GENES, COMPLETE		
MZECPP64	MAIZE CHLOROPLAST RIBOSOMAL PROTEIN 84 GENE, 1241BP		
MZECPP8P	MAIZE CHLOROPLAST RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE SUBUNIT		
MZEMTC08	MAIZE MITOCHONDRIAL APOCYTOCHROME B GENE (COB), COMPLETE CDS.		
MZEMTC03	MAIZE MITOCHONDRIAL CYTOCHROME C OXIDASE SUBUNIT I (COX I) GENE,		
MZEMTC01	MAIZE (Z. MAIZE) MT CYTOCHROME OXIDASE SUBUNIT II (COX II), 1797BP		
NEUMTC00	N. CRASSA MT CYTOCHROME OXIDASE SUBUNIT II (COX II), 1254BP		
NEUMTC01	N. CRASSA MT CYTOCHROME OXIDASE SUBUNIT 2 GENE (COX2) AND PLANKS.		
NEUMTC011	N. CRASSA MITOCHONDRIAL CL2 GENE, ATPASE SUBUNIT 4, COMPLETE CDS.		
NEUMTC03	N. CRASSA CYT-4 MUTANT MITOCHONDRIAL 18S rRNA, EXONS 1 AND 2.		
OBMTCT02	OBENTHERA BERTERIANA MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT II		
OBMTCT07	OBENTHERA MOOREI PLASTID GENE FOR PRE-APOCYTOCHROME F 1672BP		
PANMTCT01	P. ANSERIANA MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT 2 GENE (COX2),		
PANMTCT01	P. AURELIA SPECIES 4 CYTOCHROME OXIDASE SUBUNIT I AND HDI GENES.		
PANMTCT02	P. AURELIA (SPECIES 4) MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT II		
PANMTCT1A	P. AURELIA SPECIES 4 MITOCHONDRIAL PROTEIN P1 GENE, ORF5 1 AND 2.		
PANMTCT1B	P. AURELIA SPECIES 4 MITOCHONDRIAL PROTEIN P1 GENE, 712BP		
PEACPCYF	PEA (P. SATIVUM) CHLOROPLAST CYTOCHROME C GENE AND FLN2, 1248BP		
PEACPD2	PEA (P. SATIVUM) CHLOROPLAST PHOTOSYSTEM II D2 PROTEIN GENE, 1537BP		
PEAMTC01	PEA (P. SATIVUM) MT COB GENE ENCODING CYTOCHROME OXIDASE II, 2038BP		
RATMTCT01	RAT MT CYTOCHROME B GENE, PRO, THR, GLU-TRNA GENES, AND URF4.		
RATMTCT01	RAT (RATTUS) MT CYTOCHROME OXIDASE SUBUNIT II GENE, 702BP		
RATMTCT01	RAT (RATTUS) MITOCHONDRIAL GENES FOR CYTOCHROME OXIDASE		
RICMTCT02	RICE (ORYZA SATIVA) MT CYTOCHROME OXIDASE SUBUNIT II (COX II) GENE.		
SALCPP88	SINAPIS ALBA CP PRE-MR 32000 PHOTOSYSTEM II MEMBRANE PROTEIN GENE,		
SNCPF88A	SOLANUM MORGANII CHLOROPLAST PSBA GENE FOR 32 KDA PROTEIN OF 1601BP		
BOVCPF88B	BOVINE CP PSBA GENE (PHOTOSYSTEM II THYLAKOID MEMBRANE PROTEIN),		
SPCPAP	SPINACH (S. OLERACEA) CHLOROPLAST P-680 CHLOROPHYLL ALPHA 2217BP		
SPCPATB5	SPINACH (S. OLERACEA) CP COUPLING FACTOR COMPLEX (ATPASE), 2414BP		
SPCPCYT8	SPINACH PLASTID GENES FOR APOCYTOCHROME B4 AND CYTOCHROME B6F		
SPCPD02B	SPINACH CHLOROPLAST PHOTOSYSTEM II PROTEIN D2, 44 KD REACTION-		
SPCPP88A	SPINACH (S. OLERACEA) CHLOROPLAST THYLAKOID MEMBRANE PROTEIN GENE		
SPCPRUBP	SPINACH CHLOROPLAST RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE SUBUNIT		
SPCPTG4	SPINACH CHLOROPLAST 16S-TRNA AND RIBOSOMAL PROTEIN L3 AND 519		
SRGMTCT02	SORDARIUM (BIODOLIA) MITOCHONDRIAL CYTOCHROME C OXIDASE SUBUNIT I		
SRGMTCT04	SORDARIUM (MUTANT HQ) MITOCHONDRIAL CYTOCHROME C OXIDASE SUBUNIT I		
TOBPC00	N. TABACUM (VAR. BRIGHT YELLOW 4) CHLOROPLAST, COMPLETE GENOME.		
TOBPCP32	TOBACCO P1, TABACUM THYLAKOID MEMBRANE PROTEIN GENE (P32) AND		
TOBPCP88A	TOBACCO (S. DUBRENYI) CHLOROPLAST THYLAKOID MEMBRANE PROTEIN GENE		
TOBPCP88B	TOBACCO CP PUTATIVE RIBOSOMAL PROTEIN C319 (P319) AND FLN2, 480BP		
TOBPCP8P	TOBACCO CHLOROPLAST RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE SUBUNIT		
TOBPCP110	INTRON IN TOBACCO CHLOROPLAST GENE RPS110 FOR RIBOSOMAL PROTEIN S10.		
TOBPCP01	TOBACCO (S. DUBRENYI) CHLOROPLAST 16S-TRNA AND RPS 17 GENES, 604BP		
TOBPCP02	TOBACCO (S. DUBRENYI) CHLOROPLAST 16S-TRNA AND RIBOSOMAL PROTEIN L2		
TRYMTCT01	T. BRUCEI MITOCHONDRIAL MAJOR CIRCLE DNA ENCODING APOCYTOCHROME B, AND		
WHATCTATP	WHEAT CHLOROPLAST ATP SYNTHASE PROTON-TRANSLLOCATING SUBUNIT GENE.		
WHATCTATP8	WHEAT CHLOROPLAST ATP SYNTHASE CP-6 SUBUNIT I GENE, COMPLETE.		
WHATCTCYF	WHEAT CHLOROPLAST CYTOCHROME F GENE, 1114BP		
WHATCTCYB	WHEAT CHLOROPLAST GENE FOR CYTOCHROME B-559, 664BP		
WHMTCT01	WHEAT MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT II (COX II) GENE.		
WHMTCT08	WHEAT MITOCHONDRIAL APOCYTOCHROME B GENE, COMPLETE CDS, 1943BP		
XLMTCT01	X. LAMPROLUM MITOCHONDRION, COMPLETE GENOME, 16336BP		
XLMTCT02	X. LAMPROLUM MITOCHONDRIAL DNA CONTAINING THE D-LOOP, AND THE 12S rRNA.		
YBOMTAP61	YEAST (S. CEREVISIAE) MITOCHONDRIAL ATPASE SUBUNIT GENE, 1452BP		
YBOMTAT91	YEAST (S. CEREVISIAE) MITOCHONDRIAL ATPASE PROTEOLIPID (SUBUNIT 9)		
YBOMT003	YEAST (S. CEREVISIAE) MITOCHONDRIAL COX2 GENE ENCODING CYTOCHROME C		
YBOMT004	YEAST (S. CEREVISIAE) MITOCHONDRIAL CYTOCHROME C OXIDASE SUBUNIT		
YBOMT005	YEAST (S. CEREVISIAE) MITOCHONDRIAL CYTOCHROME B SHORT GENE, 629BP		
YBOMT001	YEAST (S. CEREVISIAE) MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT I		
YBOMT002	YEAST (S. CEREVISIAE) MITOCHONDRIAL CYTOCHROME OXIDASE SUBUNIT 2		
YBOMT0011	YEAST (S. CEREVISIAE) MT CL1 GENE ENCODING ATPASE SUBUNIT A, 351BP		
YBOMT00	YEAST (S. CEREVISIAE) MITOCHONDRIAL OXIDOLE INTERCISTRONIC REGION		